UNIVERSITY
OF MICHIGAN
APR 5 1955
SOCIAL SCIENCE
LIBRARY
BF
1
.J86

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Concept Identification as a Function of Irrelevant Information and Instructions: E. J. ARCHER, L. E. BOURNE, JR., AND F. G. BROWN 153

A Comparison of Reversal Shifts and Nonreversal Shifts in Human Concept Formation Behavior: H. H. KENDLER AND M. F. D'AMATO 165

Concept Formation: A Problem in Human Operant Conditioning: E. J. GREEN 175

Shock for Right and Wrong Responses During Learning and Extinction in Human Subjects: C. M. FREEBURNE AND M. SCHNEIDER 181

Discriminative Skill and Discriminative Matching in Perceptual Recognition: J. S. BRUNER, G. A. MILLER, AND C. ZIMMERMAN 187

Extinction as Due to the Changed Direction of a Psychological Force: K. JOHDAI 193

Flicker Fusion Measurements and Anxiety Level: S. GOLDSTONE 200

Constant Velocity Tracking as a Function of *S*'s Handedness and the Rate and Direction of the Target Course: D. A. GRANT AND N. F. KAESTNER 203

The Relation Between the Rate Threshold for the Perception of Movement and Luminance for Various Durations of Exposure: H. W. LEIBOWITZ 209

Dark Adaptation After Intermittent and Continuous Pre-Exposures of Equal Duration: F. A. MOTE AND E. R. ADAIR 215

A Retinal Excitation Gradient in a Uniform Area of Stimulation: L. KRUGER AND J. R. BONAME 220

American Psychological Association

Vol. 49 No. 3 March 1955

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1955 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Department of Psychology, Stanford University, Stanford, California, until June 1, 1955. After June 1 address manuscripts to the editor at Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1955, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 49, NO. 3 MARCH, 1955

CONCEPT IDENTIFICATION AS A FUNCTION OF IRRELEVANT INFORMATION AND INSTRUCTIONS [1]

E. JAMES ARCHER, LYLE E. BOURNE, JR., AND FREDERICK G. BROWN

University of Wisconsin

Except for a few studies (e.g., **5, 13**), little attempt has been made to quantify and systematically examine the variable of task complexity in concept formation and problem solving. Most previous studies have considered such variables as distribution of practice (**4, 8, 16**), previous experience (**1, 2, 7, 14**), and nature of concepts to be attained (**3, 6, 11**). Two recent theoretical papers (**12, 17**) suggest that the variable of task complexity might be quantified by using probability measures as defined by information theory (**10, 15**).

If the stimuli to be classified conceptually can be described in terms of a specific number of dimensions and levels, then the appearance of any particular stimulus can be stated in probability terms. As the number of possible stimuli increases, the amount of information contained in each alternative also increases. This amount of information can be quantified and defined as $\log_2 x$, where x is equal to the number of possible stimuli. For example, in the Oseas and Underwood study (**16**) three levels of each of three dimensions of form, size, and shade were shown in all possible combinations. Each pattern, e.g., large black triangle, contained $\log_2 27$ or 4.75 bits. Of this total amount of information, 1.58 bits were irrelevant and 3.17 bits were relevant. These conditions were obtained by having two dimensions, e.g., form and size, relevant and the other dimension, shade, irrelevant. No attempt was made in that study to vary the amount of information since distribution of practice was the major variable under consideration.

If for a particular set of stimuli, all levels of all dimensions are easily and completely distinguishable, i.e., transmit an equal amount of information, then a bit of information from one dimension should be as difficult to encode and classify as that from any other dimension. If this were true, then it would be possible to quantify task complexity independently of the particular dimensions involved. To this end a preliminary experiment was conducted, and, following favorable

[1] This research was conducted under United States Air Force Contract No. AF 33(038)-23294 between the University of Wisconsin and the 6564th Research and Development Group, Personnel and Training Research Center, Lowry Air Force Base, Denver, Colorado.

results, the present research was undertaken.

We do not at present have a mathematical model which will predict the number of responses or time to solution. However, we can reasonably expect it to be a positively accelerated exponential function since the number of possible stimuli doubles with each additional bit of information and the number of hypotheses which can be formulated follows the function:

$$h = i^2/2 + 3i/2 + 1$$

where h = number of hypotheses and i = number of bits of irrelevant information. In all cases, two bits are relevant.

General Procedure and Apparatus

Task.—During a particular session, *S* was presented with all possible stimulus combinations within the limits of that session's task complexity.

The *S* was instructed to classify each pattern into one of four categories, indicating his choice by pressing one of four buttons. Each button corresponded to a combination of the two levels and two dimensions which were relevant, e.g., large-circle would be indicated by one button and small-circle would be indicated by another. A dimension is relevant if it is necessary for the correct classification of patterns. An irrelevant dimension is defined as one which appears at each of its two levels but can never be used to correctly identify any of the four concepts. If a "dimension" was neither relevant nor irrelevant for a particular day's problem, only one level of that "dimension" would appear. For example, if the dimension of vertical direction was neither relevant nor irrelevant, all patterns would move either up or down but not in both directions.

A correction procedure was used, i.e., a new pattern appeared only after *S* made a correct response. Errors were indicated to *S* by a red panel lamp. Correct responses were indicated by a green lamp and the subsequent appearance of a new pattern. The criterion was 32 consecutive correct identifications; the attainment of this criterion was automatically signaled to *S* by a white panel lamp.

Since *S*'s push buttons bore no markings and varied in meaning from day to day, *S* had to learn the significance of each button by trial and error.

Apparatus.[2]—Aside from *S*'s control panel, which contained four push buttons and three signal lamps, there were four major units to the apparatus. The patterns were presented on a 7-in. cathode ray oscilloscope (RCA WO-56A) which remained intact except for one relay in the brightness control circuit. A signal generator provided the input to this oscilloscope through the use of two Miller integrators and several switching and time delay circuits. A Western Union tape transmitter with a continuous loop of tape provided a "random" sequence of patterns which repeated itself after 192 presentations. The recording console included a Standard Electric time clock (.01-min. scale), two Veeder-Root counters, and an automatic reset stepping relay. The clock started when the first pattern was presented and automatically stopped when *S* attained the criterion. The counters recorded correct or incorrect responses, and the stepping relay advanced with each correct response and returned to zero with each error. Automatic control of the timer clock and the criterion signal to *S* was provided by this component.

Hereafter, this apparatus will be designated as the Concept Identification Apparatus.

Experiment I

The purpose of the first study was to investigate the effect of variations of task complexity upon the acquisition of simple concepts. For the particular problems to be solved, two bits of information were relevant and one, two, or three bits were irrelevant.

Procedure

Subjects.—Twenty-four paid, male, University of Wisconsin students served for three days each in this experiment. For practically all *S*s this service was on consecutive days. Each *S* was presented with detailed tape-recorded instructions on the first day which described the nature of a concept, the operation of *S*'s response controls, and the criterion of the study. The *S* was told to work for both speed and accuracy.

Design.—Eight different 3×3 greco-latin squares were used. The rows corresponded to different *S*s, the columns to successive stages of practice, the latin letters to problems (each containing two bits of relevant information), and the greek letters to amount of irrelevant information present in the problem.

[2] The writers are indebted to David Feiges and Manfred Kahn for their technical assistance in the development of this apparatus.

From a group of three bilevelled dimensions, form, size, and brightness, two dimensions were selected as relevant to the solution of a particular day's problem. The third dimension was irrelevant. There are only three possible such combinations: form and size, form and brightness, and brightness and size, with brightness, size, and form irrelevant, respectively. The specific levels of these dimensions were: circle and ellipse (form), large and small (size), and bright and dim (brightness).

From the group of two bilevelled dimensions of horizontal direction (right to left or left to right) and speed (fast or slow), none, one, or two dimensions were selected and added to the previous combination of two relevant and one irrelevant bits. This yielded problems having three levels of complexity; all had two bits of information relevant and one, two, or three bits irrelevant.

Results

Time to criterion.—One of the basic response measures in this study was time to criterion measured in minutes. As shown in Fig. 1, the time to solution increased as the amount of irrelevant information increased from one to three bits. Figure 2 indicates that time to solution markedly decreased with practice. The significance of these and other sources of variation is shown in Table 1.

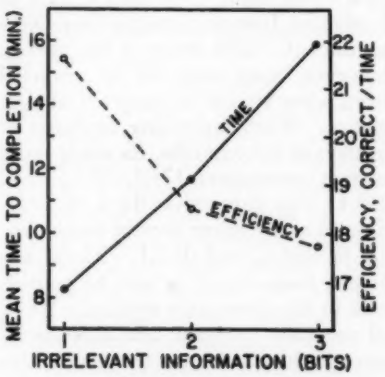

FIG. 1. The mean time in minutes to solution is plotted as a function of level of irrelevant information. The scale for this curve is the left-hand ordinate. The broken line curve shows efficiency (correct responses per minute) plotted against irrelevant information. The scale for this curve is shown on the right-hand ordinate.

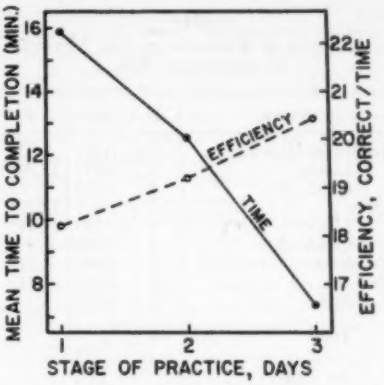

FIG. 2. The mean time in minutes to solution is plotted as a function of stage of practice (days). The scale for this curve is the left-hand ordinate. The broken line curve shows efficiency (correct responses per minute) plotted against irrelevant information. The scale for this curve is shown on the right-hand ordinate.

The differences among the mean log times to solution as a function of amount of information were significant ($F = 6.87$, 2 and 42 df, $p < .01$).[3] Since with each additional bit of irrelevant information the number of different stimuli doubled and the number of hypotheses increased according to the parabolic function of $h = i^2/2 + 3i/2 + 1$, we might reasonably expect a nonlinear increase in time to solution. A general test for nonlinearity was made using orthogonal polynomials (9). The results of such an analysis are also shown in Table 1.

[3] Because of nonnormality and heterogeneity of variance of the time-to-criterion scores in Exp. II of this paper the data were logarithmically transformed. For the sake of consistency in reporting the results of the two studies the time scores from both experiments were transformed. To be certain that information was neither lost nor spuriously introduced by these transformations an analysis of the original time scores of Exp. I was compared with an analysis of transformed scores as shown in Table 1. There were no differences in confidence levels for any of the sources of variation.

TABLE 1

ANALYSIS OF VARIANCE OF LOG MINUTES TO CRITERION: EXP. I

Source	*df*	*MS*	*F*
Irrelevant information	2	.4937	6.87*
Linear	1	.9872	13.73**
Quadratic	1	.0002	
Days	2	.8803	12.25*
Linear	1	1.7599	24.48**
Quadratic	1	.0006	
Problems	2	.1058	1.47
Ss	23	.0966	1.34
Residual	42	.0719	
Total	71		

* $p < .01$.
** $p < .001$.

For the two partitioned sums of squares for the Irrelevant Information term only the linear component was significant. The quadratic component was not significant, indicating that the hypothesis of nonlinearity was untenable.

A marked day-to-day practice effect was obtained ($F = 12.25$, $p < .01$). Although no a priori hypothesis was made as to the form of this function, we might expect that it would be negatively accelerated with an asymptote equal to the time required to make 32 key-pressing responses. Here, too, in an orthogonal polynomial analysis, only the linear component was significant.

Another *F* ratio is of particular interest at this point. There were no significant differences among the problems. The paired dimensions and their means for time to solution were: size-brightness, 10.71 min.; size-form, 12.36 min.; and brightness-form, 12.82 min. The *F* ratio for the comparison of these means was less than unity. When the analysis was based on log transformed data a nonsignificant *F* of 1.47 was obtained.

Correct and incorrect responses.—The other basic response measures which were obtained were the number of correct and incorrect responses made prior to solution. The total variance for each of these measures was analyzed as before. The mean number of errors associated with one, two, and three bits of irrelevant information were 79.5, 101.0, and 173.8, respectively. The significance of the differences among these means is indicated by the *F* ratio of 5.23 ($p < .01$, for 2 and 42 *df*). Again, contrary to theoretical expectations, the increase in errors was not an exponential function. From an orthogonal polynomial analysis only the linear component was significant.

The mean number of errors made on the successive stages of practice (Days) were 151.5, 134.3, and 68.7. The differences among these means were significant at about the .03 level. The other *F* ratio of interest is that which tests differences among problems. This ratio indicated that no particular combination of dimensions was more difficult than another.

An analysis of correct responses to criterion was also made; the results closely parallel those of the error analysis. The same two main effects of Irrelevant Information and Days were significant. The order of the means, however, may seem to be reversed from what would be expected at first glance. With increasing amounts of irrelevant information, *Ss* made more correct responses: 175.1, 213.5, and 288.4. On successive days of practice, *Ss* made fewer correct responses: 271.6, 244.5, and 147.1. These apparent inconsistencies will be considered in the discussion section. As in all previous analyses, the differences among the problems were not significant with respect to this response measure.

Efficiency.—A derived measure which seems appropriate to this type of study is that of efficiency of re-

sponding. Although the proportion of correct responses would be one such measure, the ratio of correct responses/time was chosen since this was less dependent upon number of incorrect responses. The change in this response measure as a function of irrelevant information is shown in Fig. 1. The decrease in efficiency was highly significant ($F = 5.22$, with 2 and 42 df, $p < .01$). Although a negatively accelerated function might be expected, the decrease was not significantly nonlinear. In Fig. 2, efficiency, measured as correct responses per minute, is plotted against stage of practice. The increase in efficiency was not significant.

Experiment II

In view of the extreme scores (both time to solution and number of responses) made by some *S*s in the first experiment, it appeared reasonable that instructing *S* to approach the task analytically might facilitate overall concept formation ability. Furthermore, if such facilitation did occur, it was expected to be greater for more complex tasks; i.e., there should be an interaction between task complexity and instructions.

The second experiment was designed to test the effects of instructions and to examine further the dimension of task complexity by using problems having two bits of information relevant and as many as five bits irrelevant. All dimensions appeared as relevant at least once. These changes are in contrast to the first study in which the maximum amount of irrelevant information was three bits and the relevant information was restricted to a population of three dimensions.

Procedure

Subjects.—Thirty paid, male, University of Wisconsin students served individually for six days. Of course, none of these *S*s had served in Exp. I. All *S*s served for consecutive days except for weekends. Each *S* was presented with detailed tape-recorded instructions on the first (practice) day which described the nature of a concept, the operation of *S*'s response controls, and the criterion of the study. On the first day, the simplest possible problem was presented. This involved classifying four different patterns into four categories. No information was irrelevant. The dimensions used were horizontal direction (left or right) and form (circle or ellipse). This simple task was used to give *S* an opportunity to practice the required responses and to provide some basis for forming two groups of *S*s approximately equal in concept formation ability.

Instructions.—The general instructions given to all *S*s on the first day were essentially the same as those used in the previous study.

On the second day one-half of the *S*s were simply reminded of their task and were told that although the problem had been changed from the previous day they were to solve it in a similar manner. These *S*s are hereafter designated as the nonanalytically instructed group. The remaining *S*s ($N = 15$) were given special instructions designed to encourage analytical problem solving. The procedure for discovering the relevant and irrelevant information was outlined for a sample problem as follows: If a pattern is a large, fast ellipse and the first button pressed lights the green signal lamp, *S* knows *what* is the correct response for that pattern but not necessarily *why* it is correct. It may be correct because the pattern is large and fast, large and elliptical, or fast and elliptical. The only way to find out the correct identification of the concept is to systematically check other patterns against the same button.

For example, if a large, fast circle appears and the same button as before is pressed, either of the two consequences (right or wrong signal) should indicate something to *S* about the status of the dimensions being tested. If the same button is correct again, *S* should recognize immediately that the dimension of form (circle-ellipse) is irrelevant. On the other hand, if the response is incorrect, *S* should recognize that form must be relevant and that one of the other two dimensions must be irrelevant. To determine which of these two is to be eliminated will require testing another pattern, such as a small, fast ellipse or a large, slow ellipse with the first button. Size would be irrelevant if the first pattern were correct and speed would be irrelevant if the second were correct. The task was complicated considerably, of course, since *S* could not select the order of patterns to be classified; they appeared randomly. Nonetheless *S* was told that any problem, no matter how

complex, could be solved most rapidly by analyzing his responses and their consequences in the above manner.

Design.—Six different 5 × 5 greco-latin squares were used. The rows corresponded to different *Ss*, the columns to successive stages of practice (days), the latin letters to problems (each containing two bits of relevant information), and the greek letters to amount of irrelevant information present in the problem. The entire population of seven bilevel dimensions was form (circle or ellipse), size (large or small), speed (fast or slow), brightness (bright or dim), horizontal direction (pattern moving to the left or right), vertical direction (pattern moving up or down), and contour (steady or disturbed). Both direction components were present at the same time in that all patterns moved on a diagonal inclined 45°, plus or minus, off the vertical axis. The dimension of contour (steady or disturbed) was obtained by modifying the signal input to the oscilloscope with an audio oscillator. When the signal was disturbed, the circles and ellipses would appear as coronets; when the signal was steady, the patterns appeared as smooth circles or ellipses. With seven bilevelled dimensions there are 21 combinations of dimensions taken two at a time. From this population of 21 possible problems, six were randomly selected with the restriction that each dimension be used at least once and no more than twice. Five problems were identified by latin letters in the 5 × 5 greco-latin squares, and the sixth was used as the practice problem. The five pairs of relevant dimensions were (*a*) size and form, (*b*) vertical direction and size, (*c*) speed and brightness, (*d*) horizontal direction and contour, and (*e*) vertical direction and speed. On each experimental day, *S* was presented with a different problem. The previous study indicated that the problems would probably be of equal difficulty, and the design permitted a test of this assumption. From the five dimensions remaining after the two relevant dimensions have been designated, 1, 2, 3, 4, or 5 dimensions were randomly selected to be irrelevant.

Results

Comparability of groups.—The initial equality of the two groups, which later differed in terms of instructions, was indicated by several analyses of variance of response measures taken on the first (practice) day. The *F* ratios comparing the two groups in time to solution, total responses, and efficiency (defined as correct responses per minute) were all less than unity. The correlations between each of these measures on the practice day and the total comparable scores for all five experimental days, however, were not significant.

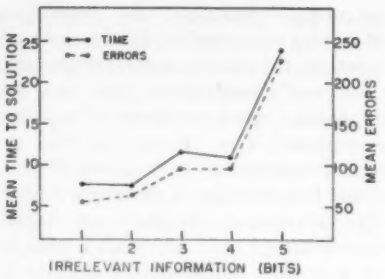

FIG. 3. Mean time to criterion and mean number of errors as a function of amount of irrelevant information as measured in bits. The vertical displacement of the two curves is meaningless since they are plotted on different scales.

Time to criterion.—The response measure of major interest is time required to attain the criterion of 32 consecutively correct identifications of the concepts. The increase in time to solution as a function of the amount of irrelevant information, as measured in bits, is shown in Fig. 3. The distribution of time scores was highly skewed with most of the scores clustered about a median of 7.5 min., but about 10% of the scores were in excess of 30 min. Because of obvious nonnormality and its attendant heterogeneity of variance, the raw time scores were logarithmically transformed. An analysis of variance of log time to criterion is shown in Table 2. The log transformation eliminated the heterogeneity of variance. As may be seen in Table 2, the main effect of instructions was negligible. The error term for evaluating the significance of the Instruction term was *Ss*/Instructions. For 1 and 28 *df*

TABLE 2

ANALYSIS OF VARIANCE OF LOG MINUTES TO CRITERION: EXP. II

Source	*df*	*MS*	*F*
Instructions	1	.1409	1.14
Ss/Instructions	28	.1235	1.25
Irrelevant information	4	.9431	9.56***
Linear	1	3.0616	31.04***
Quadratic	1	.5608	5.69*
Cubic	1	.0833	
Quartic	1	.0666	
Days	4	.4098	4.15**
Problems	4	.4785	1.81
Information × Instructions	4	.2620	2.66*
Days × Instructions	4	.0296	
Problems × Instructions	4	.2652	2.69*
Pooled residual	96	.0987	
Total	149		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

an F of 4.20 would be needed at the .05 level of confidence—an F of 1.14 was obtained. Another source of variance of immediate interest is that assigned to amount of irrelevant information. The error term for evaluating the effect of this source is the Pooled Residual. The F ratio of 9.56 indicates that the means for each of the five levels of irrelevant information deviate about a grand mean to a significant degree ($p < .001$). The sum of squares associated with the 4 *df* for irrelevant information was further analyzed through the use of orthogonal polynomials (**9**). The significant ($p < .001$) linear component indicates that the over-all slope of the curve of the five means is greater than zero. Furthermore, the significant ($p < .02$) quadratic component indicates that this increase is nonlinear. Since the higher-order terms are not significant, the function probably can best be described by a parabola. The best fitting least squares equation was found to be $y = 12.606 - 5.159x + 1.589x^2$.

Although the different instructions did not produce an over-all difference in the means of the two groups, the performance degradation of the nonanalytically oriented group had a somewhat steeper slope than that of the analytically oriented group. The significance of this performance differential as a function of instructions and task complexity is indicated by the F ratio of 2.66 ($p < .05$) for Irrelevant Information × Instructions mean square divided by the Pooled Residual term.

As would be expected, the day-to-day transfer effect of learning-to-learn concepts was highly significant ($F = 4.15$, $p < .01$). For the number of days covered by the present experiment the decrease was linear although presumably it would eventually become negatively accelerated. Another source of variation which is of interest is that associated with the different problems. Since the Problems × Instructions interaction was significantly greater than the Pooled Residual and the problems could be considered as randomly selected from a population of problems, the Problems × Instructions interaction was the appropriate error term for evaluating Problems. For 4 and 4 *df* an F ratio of 6.39 would be needed at the .05 level of confidence—an F ratio of 1.81 was obtained. The significant Problems × Instructions interaction was apparently due to the somewhat greater variability of the nonanalytically instructed group.

Correct and incorrect responses.—Although this task was self-paced, and, except for being instructed to try to solve the problem as quickly as possible, *S* was under no obligation to respond, the correlation between time to criterion and correct responses was .97. The correlations between time and errors and time and total re-

sponses were .97 and .99, respectively. In view of these remarkably high correlations, a description of variations in correct and incorrect responses as a function of the experimental variables can be brief.

A distinction must be made between making a correct response and having the correct hypothesis. The probability of making a correct response in this experiment was always .25 since one of four buttons was always correct for each pattern. The probability of having the correct hypothesis, however, varied as a positively accelerated function of the amount of irrelevant information. As would be expected, the number of correct responses increased as the amount of irrelevant information increased since *S* must make more responses to test the greater number of hypotheses. The increase in number of correct responses was significant ($F = 9.46$, $p < .01$).

A better measure of goodness of performance is the number of errors as a function of irrelevant information. The mean number of errors for each level of irrelevant information is presented in Fig. 3. The almost parallel forms of this and the time curve are not unexpected in view of the very high correlation between these measures. The vertical displacement of the two curves is meaningless since the two are plotted on different scales. A log transformation was made since the distribution of errors was highly skewed positively. Again the effect of different instructions was negligible; the *F* ratio was less than unity. The effect of increased amounts of irrelevant information was to increase the number of errors in what was probably an exponential function. From an orthogonal polynomial analysis only the linear component was found to be significant. This analysis, however, was based upon log transformed scores. A similar analysis based upon the raw error scores had a significant heterogeneity of variance. The *S*s in the nonanalytically oriented group were significantly more variable than those in the analytically oriented group ($F = 5.32$, 14 and 14 *df*, $p < .01$). Even from day to day the nonanalytically oriented group was more variable in terms of number of errors ($F = 3.29$, 48 and 48 *df*, $p < .01$).

There was a significant day-to-day decrease in the number of errors ($F = 5.46$, 4 and 96 *df*, $p < .01$). Again the Problems × Instructions interaction was significantly greater than the Residual mean square and, therefore, was used as an error estimate to test the differences between the problems. The differences between problems were not significantly greater than would be expected by chance.

Efficiency.—Although it has been demonstrated that both number of correct responses and errors increase as a function of amount of irrelevant information, there may be a differential rate of increase of these two response measures. The derived measure of the ratio of errors/correct responses was computed and an analysis of variance of these ratios was made. The day-to-day transfer effect was indicated by a significant ($F = 5.26$, 4 and 96 *df*, $p < .01$) decrease in the ratio of errors to correct responses. This response measure increased (.332, .411, .468, .551, .570) as the amount of irrelevant information increased from one to five bits, respectively. The differences among these means were significant ($F = 6.91$, 4 and 96 *df*, $p < .01$).

A final performance measure of interest is that of efficiency of responding as measured by the number of correct responses per minute. To obtain this measure the number of correct

responses on a given day was divided by the total time to attain criterion. The total variance of these ratios was analyzed and only two sources of variance proved significant. There were considerable individual differences ($F = 4.34, 28$ and $96\ df, p < .001$). The other source of significance was Irrelevant Information. Efficiency tended to decrease as indicated by the mean values of 18.20, 17.63, 17.79, 16.13, and 16.26 correct responses per minute for the five increasing levels of irrelevant information. The significance of the differences among these means was between the .05 and .01 levels of confidence ($F = 3.05$, 4 and 96 df). The analytically instructed group was numerically more efficient, but the difference between the means for the two groups of Ss was not statistically significant.

Discussion

The present experiments represent a compromise between two types of concept formation studies: The Hovland-Weiss design on the one hand and those represented by the Oseas-Underwood design on the other. In the former the information contained in a stimulus can be easily defined quantitatively. When *S* is told of the number of relevant levels and dimensions, the minimum number of positive or negative instances necessary for solution can be logically determined. A limitation of the Hovland design is that only a single concept is identified at a time and its identity is arbitrary. In studies of the Oseas-Underwood type, several concepts are to be identified, and all levels of all relevant dimensions are used. A limitation of such studies is that *S* must also learn an artificial language with which to identify the several concepts. Confounded with concept complexity is the difficulty of learning the language.

The present studies quantified the stimuli, required the identification of several concepts, and reduced the emphasis on learning an artificial language. In these experiments it was as if *S* served in four Hovland-type experiments simultaneously. Each stimulus was a positive instance for one concept identified by one button, and a negative instance for the other three concepts (buttons). Instead of learning an artificial language, *S* could designate (to himself) a button "large-fast," or "bright-circle," or any other code names he preferred.

In this type of study, negative instances as such probably do not transmit much information, partly because *S* is searching for correct responses and partly because an error means only that one of the other three buttons must be pushed. In Exp. I none of the response measures had a significant nonlinear component although nonlinearity would have been consistent with the exponential increase in hypotheses as a function of irrelevant information.

The results of Exp. II, on the other hand, provide support for the hypothesis that concept formation performance will be a positively accelerated function of the amount of irrelevant information present in the problem. Two reasons for the slight inconsistency between the results of the two experiments are (*a*) a greater range of task complexity was sampled in the second study, and (*b*) fewer restrictions were placed upon the selection of relevant dimensions. By using problems having as many as five bits of information irrelevant, a better test of the variable of task complexity was made. The selection of dimensions which were to be relevant in Exp. II was not restricted as was done in Exp. I, and therefore *S* was given no bias toward testing the relevant dimensions sooner than would be expected by chance. If such a bias did exist, the effectiveness of the irrelevant information would be attenuated. To some extent a similar bias may have been operating because *S* served for more than one session. To further eliminate this bias of choosing some dimension(s) in favor of others, a nonrepeated measures factorial design could be used. This, of course, would eliminate the day-to-day practice effect.

In general the differences between the pairs of dimensions which defined *S*'s problem for a given session were not significant. In the first experiment the *F* ratio was frequently less than unity for differences between problems. These results are interpreted to mean that at least for the dimensions of form, size, and brightness there was no order of dominance for concept formation. Any single dimension was equally effective as an inhibitor; any pair was as identifiable as the other two pairs.

The results are less clear cut in Exp. II. For three of the response measures (time to cr terion, errors, and the ratio of correct responses/minutes) the differences among problems were not significant. For two of the response measures (errors/correct responses and correct responses) the differences among the problems were significant at just beyond the .05 level of confidence. Some doubt about the significance of these differences was indicated by the lack of agreement between the two measures of errors/correct responses and number of correct responses, and the identification of the most difficult problem. As measured by time to criterion the numerical ordering of the problems in terms of increasing difficulty was: vertical direction and size, horizontal direction and contour, vertical direction and speed, size and form, and, finally, speed and brightness.

It is of interest, however, to observe that the levels of those dimensions which *probably* were more difficult to discriminate, i.e., speed, brightness, and form, were associated with the more difficult problems. Although the two forms of circle and ellipse differ on an absolute scale, they were not easily discriminated because of an artifact of the apparatus. Because the phasing circuit was unstable throughout the extent of the pattern's path, the circles were somewhat egg-shaped except in the center of the scope. Consequently, small, disturbed circles and ellipses were differentially more difficult to discriminate. The more easily discriminated levels, horizontal and vertical direction, were associated with the more rapidly identified concepts. The results indicate that task complexity could be varied in another way by holding the amount of irrelevant information constant and varying the discriminability of the levels of one or more dimensions. If the levels of a relevant dimension were indistinguishable, *S* would find the problem insoluble. The opposite effect would be obtained of course with an irrelevant dimension. If *S* were unable to discriminate between the levels of an irrelevant dimension, then no irrelevant information would be transmitted. Solution would be facilitated because there would be fewer potential hypotheses. If such results hold for all dimensions transmitting an equal amount of information, then concept formation should be independent of the specific dimensions employed. It is quite probable that in the previous studies reporting a dominance or hierarchy of dimensions (e.g., **3**, **11**) the more dominant dimensions were also transmitting more information and therefore the stimuli were not equal. The present studies seem to provide quantitative support to Dattman and Israel's "single, broad principle" (**6**). The picture of a face, for example, has many more redundant features than a holly wreath and probably transmits more information to *S*.

The apparent inconsistency of *S*'s making more correct responses with more irrelevant information can be explained as follows: Each correct response which was recorded merely indicates that the correct button was pressed (the probability of such an event was always 1/4). Such a correct response did not mean that a correct hypothesis was confirmed. The joint probabilities of a correct response *and* a correct hypothesis were 1/12, 1/24, 1/40, 1/60, and 1/84. In short, *S* had to make many more correct responses to test this increased number of hypotheses. On the first day, *S* was relatively inefficient in using correct responses to test hypotheses, but with practice he became more efficient, not in making more correct responses per minute but in needing fewer tests of hypotheses.

Although the instructions to be analy-

tical, in the second experiment, did not reduce the over-all time to criterion or number of errors, the instructions were not without effect. For all response measures except efficiency as measured by correct responses/time the variability of the analytically oriented *S*s was less than that for the control *S*s. Instructions, therefore, had a differential effect on variability. In all cases except efficiency (correct responses/time) a log transformation was necessary to eliminate the significant heterogeneity of variance. For the response measures of time to criterion and errors, the instructions had another effect—there was a significant ($p < .05$) Instructions × Irrelevant Information interaction. As had been predicted, the instructions to be analytical were more facilitating with the more complex tasks, i.e., those problems having more irrelevant information present.

It seems reasonable to predict that the effectiveness of the instructions would have been further enhanced if a noncorrection procedure had been used. In the present studies, because a self-paced correction procedure was used, little premium was attached to being right. Even early in learning *S* would always make a correct response to every stimulus. If a noncorrection procedure were used, more emphasis would be placed upon discovery of the concept and we would predict that analytically orienting instructions which suggest a technique of discovering the concept would accordingly facilitate performance.

The relationship between goodness of performance, as measured in several ways, and amount of irrelevant information is quite clear. As the amount of irrelevant information as measured in bits increases, performance degrades at a positively accelerated rate. (Although the results of Exp. I may seem to argue against this general conclusion, it is believed that the particular design which was used attenuated the effect of the irrelevant information.) The performance degradation is not simply a function of the increased number of hypotheses *S* has to test before ascertaining which dimensions are relevant and which are irrelevant. Presumably, if *S*s could utilize all of the information presented both in the stimuli and responses, time to solution would be a linear function of the number of hypotheses. Such was not the case; time to solution as a function of the number of hypotheses to be tested was also positively accelerated.

We would logically expect to be able to present problems of such complexity, e.g., ten bits irrelevant (66 hypotheses), that *S* might literally go in circles testing the same hypothesis more than once before testing them all. Extremely difficult problems, i.e., those loaded with much irrelevant information could probably be devised which would be insoluble for most *S*s.

Summary

Two experiments were conducted to test the effect in variation of amount of irrelevant information upon concept formation performance. In Exp. I 24 *S*s served for three days. Each day *S* had a different problem which was defined by two relevant dimensions from the groups of size (large or small), form (circle or ellipse), and brightness (bright or dim). The third dimension of the previous group was irrelevant. In addition, two other dimensions were also used as needed as irrelevant information. All problems had two bits of information relevant and one, two, or three bits irrelevant.

In Exp. II, 30 *S*s served for six days, to test the effectiveness of instructing *S*s to be analytical in solving concept formation problems, 15 *S*s were given no additional instructions after the first (practice) day, and 15 *S*s were given instructions on how to test a dimension for its relevance. Each day *S* was presented with a different problem having two bits of information relevant and from one to five bits irrelevant.

A major difference in the design of the two studies was that in Exp. I each dimension was relevant equally often and the number of such dimensions was restricted to three, whereas in Exp. II the only limitation which was made was that each of seven dimensions be relevant at least once during the five experimental days. The *S*s classified the patterns, which were presented on a cathode ray oscilloscope, into four categories corresponding to the combinations of two levels of two dimensions. They indicated their choices by pressing one of four buttons. A correction procedure was employed and the task was self-paced. The order of presentation of the particular problems and the amounts of

irrelevant information were determined by the literal variables of greco-latin squares. The essential results may be summarized as follows:

1. The special instructions in Exp. II did not increase the over-all performance of the analytically oriented group. The instructions, however, did reduce variability and facilitated performance on the more complex tasks.

2. Performance in concept identification degraded as a positive exponential function of amount of irrelevant information measured in bits.

3. In general, the problems, defined as the four combinations of two levels of each of two relevant dimensions, were of nearly equal difficulty. It was suggested that a critical determinant of ease of concept identification is the discriminability of the levels of the dimensions involved.

References

1. Adamson, R. E. Functional fixedness as related to problem solving: a repetition of three experiments. *J. exp. Psychol.*, 1952, **44**, 288–291.
2. Adamson, R. E., & Taylor, D. W. Functional fixedness as related to elapsed time and to set. *J. exp. Psychol.*, 1954, **47**, 122–126.
3. Baum, M. H. Simple concept learning as a function of intralist generalization. *J. exp. Psychol.*, 1954, **47**, 89–94.
4. Cook, T. W. Massed and distributed practice in puzzle solving. *Psychol. Rev.*, 1934, **41**, 330–355.
5. Cook, T. W. Amount of material and difficulty of problem solving: II. The disc transfer problem. *J. exp. Psychol.*, 1937, **20**, 288–296.
6. Dattman, P. E., & Israel, H. E. The order of dominance among conceptual capacities: an experimental test of Heidbreder's hypothesis. *J. Psychol.*, 1951, **31**, 147–160.
7. Duncker, K. On problem-solving. *Psychol. Monogr.*, 1945, **58**, No. 5 (Whole No. 270).
8. Ericksen, S. C. Variability of attack in massed and distributed practice. *J. exp. Psychol.*, 1942, **31**, 339–345.
9. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver & Boyd, 1953.
10. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
11. Heidbreder, E. The attainment of concepts: II. The problem. *J. gen. Psychol.*, 1946, **35**, 191–223.
12. Hovland, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **59**, 461–472.
13. Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.
14. Luchins, A. S. Mechanization in problem solving—the effect of *Einstellung*. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
15. Miller, G. A. What is information measurement? *Amer. Psychologist*, 1953, **8**, 3–11.
16. Oseas, L., & Underwood, B. J. Studies of distributed practice: V. Learning and retention of concepts. *J. exp. Psychol.*, 1952, **43**, 143–148.
17. Underwood, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59**, 209–220.

(Received June 21, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

A COMPARISON OF REVERSAL SHIFTS AND NONREVERSAL SHIFTS IN HUMAN CONCEPT FORMATION BEHAVIOR [1]

HOWARD H. KENDLER AND MAY F. D'AMATO

New York University

The purpose of this study was to evaluate a theoretical analysis of human concept formation behavior in a conventional card-sorting test. This analysis, which resulted from a previous investigation (**4**), assumed that card-sorting behavior on any one trial consisted of a sequence of two successive S-R associations. According to this formulation, the stimulus component of the first association would represent the test cards while the response would refer to implicit verbal or symbolic responses made to them. The stimulus of the second association would represent the cue produced by the preceding implicit response while the response would be the overt card-sorting behavior. The specific hypothesis tested in this study was that the presence of the "appropriate" symbolic cues (the stimulus of the second association), even though they might be connected to the "wrong" sorting response, would facilitate concept formation.

This hypothesis was applied to a situation involving the learning of successive concepts in a multiple solution card-sorting problem. For one group (reversal shift) the second concept to be learned was the reverse of the first concept, i.e., the cues were reversed so that *S* was required to sort the cards in a fashion opposite to that demanded initially. For the other group (nonreversal shift) the second concept was unrelated to the first concept in the sense that the basis of the correct sorting was shifted from one stimulus dimension to an entirely different one.

We would predict that the reversal shift should occur more rapidly than the nonreversal shift. This would follow from our analysis because at the completion of the learning of the first concept the implicit cues appropriate to the second concept would be present for *S*s in the reversal group; they would merely be connected to the "wrong" sorting response.

At the time this investigation was initiated Buss (**3**) reported data consistent with our prediction. In a relatively simple discrimination situation involving wood blocks which differed from each other in height, form, color, or area, he trained two groups of college students on successive discrimination problems. Group 1 learned a height discrimination and then a reverse height discrimination (reversal shift). Group 2 learned a form discrimination and then a height discrimination (nonreversal shift). The reversal shift was found to occur significantly faster than the nonreversal shift. Buss attributed this result to the fact that in the learning of the second concept, Group 2 received fortuitous partial reinforcement of the form discrimination. For example, if *S*s in Group 2 persisted in responding to the form concept during the second discrimination problem in which height was the appropriate dimension, they nevertheless would be correct 50% of the time. This partial reinforcement, according to Buss, served to maintain the form discrimination and hence retard the learning of the height discrimination by Group 2. Group 1, on the other hand, received no reinforcement of the discrimination learned previously while learning their second concept, i.e., their previously correct responses were 100% *nonreinforced*.

The present paper reports a series of three experiments. Our first experi-

[1] These studies were performed in connection with Project NR 150-064, under Contract NONR-187 (00), between New York University and the Office of Naval Research. The authors are indebted to Dr. Tracy S. Kendler for helpful suggestions in the preparation of the manuscript.

ment sought to discover whether a reversal shift would be more rapid than a nonreversal shift in an experimental situation in which the effects of partial reinforcement are eliminated. According to our analysis the control of partial reinforcement effects should not affect the superiority of a reversal over a nonreversal shift since the important factor is the presence of the appropriate verbal cues for the reversal group when the learning of the second concept is initiated. This condition should not be changed by the elimination of the effects of partial reinforcement for the nonreversal group. Our second and third experiments involved modifications of the experimental design used in Exp. I in order to discover the effect of some specific experimental techniques on the relationship between a reversal and nonreversal shift.

Experiment I

Method

Subjects.—The *S*s were 112 male students from the introductory psychology course at University Heights College of Arts and Science of New York University. A total of 16 of these *S*s were eliminated; 9 failed to meet certain requirements of the experimental design (e.g., color blindness, inability to solve the first concept, etc.) while the remaining 7 were eliminated for purposes of equating groups on their performance preceding the critical test trials. If all these 7 *S*s were included in our final results the level of significance of the differences comparing the relative effectiveness of reversal and nonreversal *S*s would all be the same. However, some of the differences between groups preceding the critical test trials would have *approached* significance. In order to be confident that our major results were a function of our experimental variables and not possibly due to sampling, we eliminated those *S*s from the various groups whose very high or very low performance prior to the critical test trials resulted in differences approaching statistical significance between to-be-compared groups.

Material.—The N.Y.U. Card Sorting Test, Form 2, was utilized. This test is patterned after the Wisconsin Card Sorting Test (1), differing mainly in that it permits the testing of more abstract and complex concepts. In this experiment two stimulus cards were used: a large orange diamond and a small dark gray ellipse with pointed ends similar in shape to a football. The form and color of the stimulus cards differed from those used in the response deck to avoid having any response card be a duplicate of either stimulus card. The deck consisted of 32 response cards which varied in terms of four forms (circle, crescent, square, and a [-shaped figure), four "colors" (black, medium gray, yellow, and red), and two sizes. The designs appeared on 3-in. square cards with the designs themselves being either 1 in. or 2 in. in height.

The two major concepts used in this experiment were shape (S) and color (C). The shape concept required rectilinear shapes (square and [-figure) to be sorted below the large orange diamond, and curvilinear shapes (circles and crescents) to be sorted below the small gray ellipse. The color concept required achromatic figures (gray and black) to be sorted below the small gray ellipse, and chromatic figures (red and yellow) to be sorted below the large orange diamond. Our experimental procedure also included the learning of a reverse shape (RS) and a reverse color (RC) concept. These reverse concepts required *S*s to sort the cards in a manner opposite to that required to learn the "direct" concept, e.g., in the learning of the reverse color concept *S* had to sort achromatic response cards below the large orange diamond, and chromatic response cards beneath the small gray ellipse.

Design.—The main purpose of this experiment was to compare the relative effectiveness of a reversal shift as compared with a nonreversal shift. This could be accomplished by the 2×2 factorial design in which half of the *S*s learned initially the shape concept and the other half learned the color concept, then half of the *S*s in each of these groups learned a reverse shape concept and the other half of each group learned the reverse color concept. Thus four groups were formed: Shape-Reverse Shape (S-RS); Shape-Reverse Color (S-RC), Color-Reverse Shape (C-RS), and Color-Reverse Color (C-RC).

In a design such as this, *S*s in the nonreversal groups (S-RC and C-RS) during the learning of the second concept would receive fortuitous partial reinforcement of sorting responses associated with their first concept. For example, placing a yellow crescent below the large orange diamond stimulus card would be a correct response for *S* in Group C-RS during the learning of the first concept. It would also be a correct response for the same *S* during the learning of the second concept. In fact, all rectilinear achro-

matic and curvilinear chromatic cards would provide partial reinforcement effects for the nonreversal *S*s during their learning of the second concept. Those cards were therefore eliminated during the *initial stage* of learning the second concept. This resulted in *S*s of both the reversal and nonreversal groups receiving 100% nonreinforcement of their *sorting responses* which had been correct for the first concept. The elimination of these aforementioned cards had also other consequences. Only rectilinear chromatic and curvilinear achromatic cards were left in the response deck during the first stage of the second concept. Since these cards resembled one of the two stimulus cards (orange diamond and gray ellipse) *both in terms of shape and color*, they would be appropriately sorted below the *same* stimulus card for both the reverse shape and reverse color concept. That is, during this first stage, the sorting responses of the response cards in the abridged deck would be *identical* for both reverse shape and reverse color concepts. This point is represented in Table 1 by the abbreviation R(SC) which indicates that in the first stage of the second concept both reverse shape *and* reverse color sorting responses are correct for all *S*s.

This condition necessitated the reinsertion of the discarded cards during the second stage of learning the second concept. In order to compare the relative effectiveness of a reversal and nonreversal shift, it was necessary to have response cards which required *different sorting responses* for the reverse color and reverse shape concept.

In order to discover whether the transfer effects resulting from a reversal or nonreversal shift were positive or negative, a control group (Group O) was added that did not receive any training on the first concept. The first exposure these *S*s had to the experimental routine involved the first stage of learning the second concept (O-SC). The pattern of reinforcement to which the control group was subjected was identical to that received by the two major experimental groups.

To summarize, our design which is schematically represented in Table 1 was an extension of the 2×2 factorial design described at the introduction to this section. The extension was necessitated by the needs (*a*) to equate partial reinforcement effects for the reversal and nonreversal shifts and (*b*) to discover the nature (positive or negative) of the transfer effects produced by reversal and nonreversal shifts. The first was accomplished by using an abridged response card deck during the first stage of the learning of the second concept which at one and the same time provided 100% nonreinforcement of the sorting response initially learned by the reversal and nonreversal groups and which also did not provide any differential reinforcement of the sorting responses associated with the learning of the reverse shape or reverse color concept. The second need was accomplished by adding a control group whose training began with the first stage of the second concept.

TABLE 1

SCHEMATIC REPRESENTATION OF DESIGN USED IN EXP. I

First Concept: Shape or Color	Second Concept: First Stage: Reverse Shape *and* Reverse Color	Second Concept: Second Stage: Reverse Shape *or* Reverse Color
Group S	Group S-R(SC)	Group S-RS
		Group S-RC
Group C	Group C-R(SC)	Group C-RS
		Group C-RC
	Group O-R(SC)	Group O-RS
		Group O-RC

Procedure.—The mechanics of the experimental procedure involved a display panel for each stimulus card below which a small cross appeared. At the start of the experiment the following instructions were orally presented by *E*:

"You will be shown a series of cards one at a time. Some of these cards will belong with this card (*E* points to the first stimulus card) and others with this one (*E* points to the second stimulus card). You must decide, for each of the cards that I shall show you, with which of these two cards (*E* points to the stimulus cards) it belongs. Indicate your choice by taking this pointer and gently tapping the mark which lies beneath the card of your choice. If your choice is correct I shall say 'Right,' but if it is incorrect I shall say 'Wrong.' Do you have any questions?"

The transition between the various stages of the experiment was made without informing *S*s of a modification in either pattern of scoring or total number of response cards that were to be used. The *S*s in all the groups were required to reach a criterion of 15 successive correct responses in their learning of the first concept and during the second stage of the learning of the second concept. The criterion during the first stage of the second concept was ten successive correct responses.

The *S*s had a limit of five decks (160 cards) to reach the criterion of learning the first concept

TABLE 2

NUMBER OF TRIALS REQUIRED TO LEARN THE FIRST CONCEPT IN EXP. I

Group	Mean	*SD*	Mdn.	Range
S-RS	2.6	3.5	0	0–11
S-RC	2.4	5.6	0	0–23
Total S	2.5	4.7	0	
C-RC	11.3	28.1	3	0–116
C-RS	8.6	17.3	3	0–72
Total C	9.9	23.3	3	

and for the second stage of learning the second concept. An *S* who during the second stage of the second concept failed to solve the problem within the limit of five decks was arbitrarily assigned a score of 160. When the 16-card deck was used during the first stage of the second concept, a maximum of 6 decks (96 cards) was permitted. The response deck during all stages of the experiment was presented in an order which was designed to preclude sequential learning.

Results

The response measure used during all stages of the experiment was the number of trials, excluding the criterion trials required to achieve the performance criterion. Table 2 reports the data associated with the learning of the first concept. The data are presented both in terms of the performance of the total number of *S*s in both Groups C and S as well as their respective subgroups. The results were highly skewed and therefore the nonparametric *U* test (6) was used to evaluate the differences.[2] The difference between the total shape and total color groups was significant at the .05 level. The fact that the median score for the learning of the shape concept was zero would suggest that this concept was, for most *S*s, the dominant initial response to the experimental situation. The difference between the two subgroups that were to be compared in the learning of the second concept (Groups C-RC and C-RS and Groups S-RS and S-RC) was, as a result of our equating procedure, slight and insignificant.

Table 3 presents the data of the first stage of the learning of the second concept. Again, we find as a result of our equating procedure slight and insignificant differences between the groups that are to be compared in the learning of the second stage of the second concept. We do find a significant difference ($p < .01$) between the combined performance of the four groups of *S*s who had learned the first concept and those groups (O-RS and O-RC) who had not, i.e., the acquisition of the first concept resulted in positive transfer effects to the first stage of learning the second concept.

The first set of results comparing the relative effectiveness of a reversal and a nonreversal shift is presented in Table 4, which reports the data associated with the learning during the second stage of the second concept.

There were two possible sets of comparisons to be made: (*a*) between groups having the same first concept but different second concepts (e.g., S-RS and S-RC); and (*b*) between

TABLE 3

NUMBER OF TRIALS TO REACH CRITERION OF LEARNING DURING THE FIRST STAGE OF SECOND CONCEPT IN EXP. I

Group	Mean	*SD*	Mdn.	Range
S-RS	6.0	4.4	4	2–16
S-RC	4.1	4.0	3	1–18
C-RC	5.4	6.2	3	1–26
C-RS	4.9	3.5	4	1–16
O-RS	12.3	13.1	7	3–50
O-RC	18.0	24.1	7	2–91

[2] Because all our results were highly skewed, the *U* test was used for all comparisons reported. The probabilities obtained with this test were doubled so that the probabilities reported in this study are consistent with a two-tailed test.

TABLE 4

NUMBER OF TRIALS TO REACH CRITERION OF LEARNING DURING THE SECOND STAGE OF THE SECOND CONCEPT IN EXP. I

Group	Mean	*SD*	Mdn.	Range	*p*
S-RS	6.7	25.6	0	0–106	<.01
S-RC	50.4	57.7	20	0–160	
C-RC	3.1	8.5	0	0–33	<.01
C-RS	60.3	52.3	40	3–160	
O-RS	16.2	39.3	0	0–160	
O-RC	29.4	45.5	6	0–160	

groups having different first concepts but the same second concept (e.g., S-RS and C-RS). Because we equated the groups in terms of their performance (prior to the beginning of the second stage of the second concept), we decided to report only the first type of comparison. The other set of comparisons led essentially to the same conclusions.

The results indicated that in *both* comparisons (Groups S-RS and S-RC and Groups C-RC and C-RS) a reversal shift occurred at a significantly faster rate than did the nonreversal shift. Since our experimental design controlled for any inherent difficulty in either of the two concepts (RC or RS) that were learned, the results are therefore consistent with our theoretical expectations.

The comparisons involving the two control groups with the various experimental groups provide information as to whether reversal and nonreversal shifts result in positive or negative transfer effects.

The comparisons between Groups O-RC and S-RC and Groups O-RS and C-RS suggest that a nonreversal shift in our experimental situation produces negative transfer effects. The former difference was significant at the .09 level whereas the latter was significant at the .0004 level. These two probabilities when combined by means of a chi-square test result in an over-all significance of <.01. Similar comparisons between the reversal groups and the control groups result in individual significances of .22 (Groups S-RS and O-RS) and .04 (Groups C-RC and O-RC) and an over-all significance of .05. These comparisons suggest that a reversal shift produces positive transfer effects and hence is consistent with our theoretical prediction.

Discussion

The present experiment was designed to compare two different formulations of human concept formation behavior. It will be recalled that Buss attributed the superiority of his reversal *S*s to the assumption that the initially learned sorting response of the nonreversal *S*s tended to be *maintained* by the partial reinforcement effects occurring during the learning of the second concept. We, on the other hand, assumed that a reversal shift should occur at a more rapid rate than a nonreversal shift because at the completion of the learning of the first concept, the symbolic cues appropriate to the second concept would be present for the *S*s in the reversal group; they would merely be connected to the "wrong" sorting response.

The results obtained during this critical second stage demonstrated the superiority of a reversal shift over a nonreversal shift and suggested that a reversal shift results in positive transfer effects while a nonreversal shift produces negative transfer effects. We are, therefore, led to conclude (*a*) that the superiority of a reversal shift over a nonreversal shift is independent of partial reinforcement effects, and (*b*) that human concept formation behavior is more adequately explained by an S-R formulation which involves a mediational process than by one which does not.

It might be argued that since it was necessary to use the entire 32-card response deck in order to evaluate the

difference between a reversal and nonreversal shift, the nonreversal group would of necessity receive partial reinforcement effects of the concept it initially learned. Consequently, the argument would run, the obtained difference between the reversal and nonreversal shift could be attributed to partial reinforcement effects during the learning of the second concept.

There are two limitations to such an analysis. (*a*) It should be remembered that in Buss' analysis the point was made that the partial reinforcement effects *maintained* the initially acquired sorting responses. Since these *sorting responses* were completely abandoned during the extinction series, it would be impossible for partial reinforcement effects to *maintain* them. (*b*) It should be recognized that this line of reasoning credits the superiority of the reversal shift to the negative transfer effects involved in a nonreversal shift. Our results suggest that the reversal shift groups learned the second concept more rapidly than did the control groups. The problem stemming from this experiment is not only to explain why the learning of the nonreversal *S*s was slowed down, but also why the learning of the reversal *S*s was speeded up.

Another possible criticism of our experimental procedure would be that since the first stage required *S*s to sort the response cards in a manner opposite to that initially learned, it forced the experimental *S*s, whether they were in the reversal or nonreversal groups, to respond in a "reversal manner" during the first stage of learning the second concept. This would result in giving *S*s in reversal groups a "head start" in the learning of their second concept. But what does it mean when it is said that the first stage forced the experimental *S*s to respond in a "reversal manner"? The abridged deck used during the first stage permitted the same sorting responses for both the reverse shape and reverse color concept. There was, however, no cue in the experimental situation which favored one concept over the other. Our position is that the symbolic cue persisting from the training on the first concept became associated with the new correct sorting response of the first stage of the second concept. In this sense the *S*s were forced to respond in a "reversal manner" because the symbolic cue from the previous training was the cue needed for a reversal shift but not the symbolic cue needed for a nonreversal shift. This analysis of the *processes underlying the card-sorting behavior* during the first stage is of course consistent with the point we sought to demonstrate, viz., the superiority of a reversal shift over a nonreversal shift is due to the availability of the appropriate symbolic cue and not due to the partial reinforcement effects maintaining the sorting responses initially correct for the nonreversal *S*s.

Although our formulation involves a mediational process it should be noted that it is defined in terms of *experimental operations* and is not dependent upon introspective reports.

The superiority of a reversal shift over a nonreversal shift as demonstrated in Exp. I might conceivably be due to some of our procedural techniques rather than to the formulation which initiated the research. It can be argued that the obtained superiority of the reversal *S*s was due to the fewer "shifts" they were required to make. Whereas the reversal *S*s had merely to reverse their original concept, the nonreversal *S*s had to "shift" to a concept based on a different stimulus dimension and then "shift" to the reverse of the new concept. The question arises as to whether the relationship reported in Exp. I would continue to hold in a situation in which the nonreversal *S*s were shifted to a concept which did not involve sorting the response cards in opposition to the relevant attributes of the stimulus card. Experiment II seeks to answer this question.

The possibility also exists that the superiority of a reversal shift over a nonreversal shift reported in Exp. I might be solely a function of the technique we used to equate partial reinforcement effects. Buss (3), of course, reported that a reversal shift was more rapid in a situation in which no attempt was made to equate partial reinforcement. But his

experiment and Exp. I differed in several respects; one being that his involved a simple discrimination situation while ours involved a conventional card-sorting test. It would, therefore, seem prudent to control for the possibility that our results were primarily due to the technique for equating partial reinforcement effects. This was done in Exp. III.

Experiment II

Method

Subjects.—The *Ss* were 62 male college students. Of these 13 were eliminated, 10 because of their inability to solve the first concept, two due to errors of the experimenter, and one to equate the groups. The significance of our major results would not have been affected by including this one discard.

Procedure.—The entire procedure, including the experimental design, was similar to the one used in Exp. I with the exception that no stimulus cards were used and the instructions were modified accordingly. The absence of stimulus cards made the second concept for the nonreversal *Ss* a "direct" concept since it did not have to be sorted in opposition to any environmental cues or to a previous mode of responding. Groups S and C each had a total of 16 *Ss* while Group O included 17 *Ss*. Half of Groups S and C received a reversal shift (Groups S-RS and C-RC), while the remaining half received a nonreversal shift (Groups S-C and C-S). Nine *Ss* in the control group learned the shape concept (Group O-S) while the remaining 8 *Ss* learned the color concept (Group O-C).

Results

There was no difference between the to-be-compared groups in the learning of the first concept. The results, when compared with the findings of Exp. I, did indicate that the learning of these concepts was appreciably slower due to the absence of the stimulus cards. The median number of trials to learn the shape concept was 32 whereas the equivalent figure for the color concept was 13.

Table 5 reports the data of the first stage of learning the second concept. Differences between the to-be-compared groups were not significant.

TABLE 5

Number of Trials to Reach Criterion of Learning During the First Stage of Second Concept in Exp. II

Group	Mean	*SD*	Mdn.	Range
S-RS	6.00	7.55	2	1–24
S-C	4.13	2.42	3.5	1–8
C-RC	1.88	.33	2	1–2
C-S	2.63	1.99	2	1–6
O-S	10.75	12.43	7	1–41
O-C	7.56	5.03	9	2–16

The performance of the four experimental groups combined was significantly superior ($p = <.01$) to that of the control group, thus suggesting positive transfer effects from the learning of the first concept to the initial stage of learning the second concept.

Table 6, which reports the major results of Exp. II, reveals that the reversal group (C-RC and S-RS) reached the criterion during the second stage of learning the second concept in significantly fewer trials than the corresponding nonreversal groups (C-S and S-C). It should be noted that the transition from the first to the second stage required zero trials for both reversal groups, thus suggesting that the cue resulting from the symbolic response appropriate to the

TABLE 6

Number of Trials to Reach Criterion of Learning During the Second Stage of the Second Concept in Exp. II

Group	Mean	*SD*	Mdn.	Range	p
S-RS	0.0	0.0	0	0–0	<.01
S-C	25.6	23.1	18.5	4–75	
C-RC	0.0	0.0	0	0–0	<.01
C-S	78.1	54.7	78	5–160	
O-S	9.4	23.0	0	0–70	
O-C	31.6	48.4	17	0–160	

first concept tended to become connected to the correct sorting response during the initial stage of the second concept.

A comparison between the reversal groups and their respective control groups reveals a difference ($p = .02$) in favor of Group C-RC over Group O-C and a difference ($p = .40$) in favor of Group S-RS over Group O-S. The two probabilities combined by means of a chi-square test result in an over-all significance level of .05. The comparisons between the nonreversal groups and their respective control groups indicate that Group O-S was superior to Group C-S at the .01 level whereas Group O-C was superior to Group S-C at the .50 level. These two probabilities result in a combined .02 over-all level of significance. These results are consistent with those obtained in Exp. I in suggesting that when compared with no training on the first concept, a reversal shift leads to positive transfer while a nonreversal shift results in negative transfer.

Experiment III

Method

Subjects.—The *Ss* were 58 male undergraduate students. Of these, six were eliminated for failure to satisfy the requirements of the experimental procedure (five failed to solve their first concept and one was color blind) and four were eliminated for purposes of equating the groups on their performance during the learning of the first concept. The inclusion of these discarded *Ss* would not have affected the significance of the results reported between the various groups in their learning of the second concept.

Procedure.—The experimental procedure duplicated that of Exp. II except that the first stage of learning the second concept was omitted in the present experiment. The *Ss* in this experiment went directly to the second concept with no attempt made to equate partial reinforcement effects between the reversal and nonreversal groups.

TABLE 7

Number of Trials to Learn the Second Concept in Exp. III

Group	Mean	*SD*	Mdn.	Range	*p*
S-RS	5.8	6.3	2.5	1–20	<.01
S-C	42.5	48.2	20	9–160	
C-RC	2.8	2.1	2	1–8	<.01
C-S	60.4	43.4	49	8–135	
O-S	25.1	22.6	20	5–83	
O-C	29.1	20.4	22	7–57	

Results

The difference between the to-be-compared reversal and nonreversal groups in the learning of the first concept was insignificant. The median number of trials to learn the shape concept was 17 while the equivalent figure for the color concept was 36.5. The same relationship of a superiority of reversal over nonreversal shift, which was found in the first two experiments, prevailed also in Exp. III. These data are reported in Table 7. The comparisons between Groups C-RC and C-S and Groups S-RS and S-C were both significant at beyond the .01 level. The comparisons between the reversal groups and their corresponding control groups (Groups C-RC and O-C and Groups S-RS and O-S) were both significant beyond the .01 level. Corresponding comparisons between nonreversal and control groups yielded insignificant differences although in a direction consistent with negative transfer.

Discussion

The three experiments stemmed from a theoretical viewpoint which assumed that concept formation behavior in a card-sorting situation consisted of two successive S-R connections. The stimulus component of the first association, according to our view, represented the

test cards while the response referred to the implicit responses made to them. The stimulus of the second association represented the cue produced by the preceding implicit response while the response was the overt card-sorting behavior. This assumption led us to predict that the presence of "appropriate" implicit stimuli would facilitate the learning of a concept which required merely the change from one to another of two possible sorting responses. This prediction was borne out by the results of all three experiments which showed a reversal shift was accomplished more rapidly than a nonreversal shift.

As partial reinforcement effects were equated during the first and second experiments, it seems unlikely that Buss' hypothesis (3), which attributed a similar finding to the retarding effects of partial reinforcement fortuitously received by nonreversal *S*s during their learning of the second concept, can account for our results.

The results of Exp. III in which partial reinforcement effects were not equated suggested that our experimental technique for equating partial reinforcement was not mainly responsible for the obtained superiority of the reversal shift.

The second and third experiments, which did not involve the use of stimulus cards, suggested that the superiority of a reversal shift over a nonreversal shift was independent of whether the shift was to a reverse concept involving stimulus cards or to a direct concept involving no stimulus cards.

All three experiments consistently demonstrated that a reversal shift results in positive transfer effects. This conclusion is based upon the superior performance of the reversal groups in the acquisition of their second concept when compared to the control group. This result was consistent with our theoretical expectations. It is of interest to note here that Buss reported negative transfer effects for his reversal *S*s. In his experiment, however, the *S*s were presented with a total of only 12 stimuli during the "learning" of the first concept. Since his reversal group responded correctly only at a 58% level of accuracy while "learning" the first concept, it is unlikely that many of his *S*s were making the appropriate symbolic response prior to their reversal shift.

The initial two experiments suggested that a nonreversal shift produced negative transfer effects while the third experiment failed to provide significant negative transfer effects although the results tended to be in that direction.

In closing it may be appropriate to mention a point associated with the problem of extending learning theory to human problem-solving behavior. It has been demonstrated (2) that a reversal shift in a simple discrimination situation with animals produces negative transfer effects. The fact that the opposite relationship was reported in this experiment does not preclude the extension of learning theory to problem-solving behavior. Being able to extend a theory from one empirical area to another does not mean that the same empirical relationships will hold; it means, rather, that the same theoretical principles will operate. Whether or not learning theory can be extended to human problem-solving behavior is a problem for the future to decide. It is our belief that chances for success will be increased by emphasizing the importance of response-produced stimuli as has been done in our theoretical analysis.

Summary

Three experiments were designed to evaluate an analysis of card-sorting behavior which assumed that such behavior on any one trial consisted of a sequence of two successive S-R associations. According to this formulation, the stimulus component of the first association would represent the test cards while the response would refer to the symbolic response made to them and the stimulus of the second association would represent the cue produced by the preceding symbolic response, while the response would be the overt card-sorting behavior. The specific hypothesis tested was that the presence of the appropriate symbolic cues, even though they might be connected to the wrong sorting response, would facilitate concept formation.

According to this hypothesis a reversal shift should occur at a more rapid rate than a nonreversal shift because at the completion of the learning of the first concept, the symbolic cues appropriate to the second concept would be present for the *S*s in the reversal group; they would merely be connected to the "wrong" sorting response.

The experimental results obtained in a situation involving the learning of successive concepts in a multiple solution card-sorting problem were consistent with our expectations. The superiority of a reversal shift over a nonreversal shift was discovered to be independent of (*a*) the presence or absence of any partial reinforcement effects from previously learned concepts, and (*b*) whether the concept shifted to was a reverse or a "direct" concept. The results also indicated that a reversal shift produces positive transfer effects.

References

1. Berg, E. A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15–22.
2. Blum, R. A., & Blum, J. S. Factual issues in the "continuity" controversy. *Psychol. Rev.*, 1949, **56**, 33–50.
3. Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.
4. Kendler, H. H., & Vineberg, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, **48**, 252–259.
5. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

(Received April 26, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

CONCEPT FORMATION: A PROBLEM IN HUMAN OPERANT CONDITIONING [1]

EDWARD J. GREEN

Harvard University[2]

Vinacke has remarked on the paucity of techniques used in the study of concept formation in humans (6), and Leeper has pointed out that grouping animal discrimination learning with concept formation might result in fruitful interchanges between fields (3). The present study was designed to apply operant conditioning techniques to the experimental analysis of human behavior. Specifically, this experiment was aimed at determining how a complex discrimination, or concept formation, is controlled by fixed-ratio schedules of reinforcement (4) and by the length of time that discriminative stimuli are presented to *S*s.

METHOD

Response.—Two responses were studied. One was tapping a telegraph key; the other was a verbal statement by *S* of what seemed to him to be the proper occasion for tapping the key. The proper occasions were particular discriminative stimuli in the presence of which tapping the key resulted in reinforcement during conditioning. Reinforcement consisted of crediting *S* with points on a counter for tapping the key in the presence of a particular class of stimulus patterns.

Experimental plan.—Six experimental and three control groups of five *S*s each were used. Three fixed ratios of reinforcement were used. These were 1:1 (continuous), 15:1, and 30:1 responses to reinforcement. Discriminative stimuli were presented to *S*s for 3 sec., 30 sec., and 60 sec., depending upon the experimental group. Since it is difficult or impossible to maintain or even establish a 30:1 schedule in a 30-sec. interval, the use of this schedule was limited to the 60-sec. group. The same was true of this schedule and the 15:1 ratio schedule in the 3-sec. group where only one response per stimulus was permitted. Both conditioning and extinction were studied.

Apparatus.—The *S* was seated before a panel that separated him from *E*. Stimulus cards were presented one at a time through an opening in the panel. Reinforcements were given on one of the three schedules for tapping the key. No change in procedure was evident to *S* when the extinction trials began other than the fact that he received no more points for tapping the key.

Discriminative stimuli.—Figure 1 shows six of the discriminative stimulus patterns used. Patterns can be made more or less alike by blacking-in the same or different circles. In this experiment, cards that were occasions for reinforcement (S^D's) were made as similar to those which were not (S^Δ's) as possible. The two stimulus classes differed only in that the three top circles of the S^D's were always blacked-in. These were circles *a*, *b*, and *c* in Fig. 1. The S^Δ's might have none, one, or two of these circles blacked-in, but never all three. For every S^D with a particular pattern in the remaining six circles, there was an S^Δ with an identical pattern in those circles. This characteristic of the two stimulus classes is also indicated in Fig. 1. No two cards, either S^D's or S^Δ's had an identical over-all pattern. This condition was imposed to insure that *S*s discriminated classes of stimuli rather than particular instances.

Twenty cards were shown in conditioning and 20 in extinction. Before each *S* was tested, the cards were shuffled ten times. Half of each set of 20 cards were S^D's and half S^Δ's, but the first card of both the conditioning and the extinction series was an S^D. This was done in extinction to avoid merely extending the conditioning series an additional interval, since S^Δ's are never occasions for reinforcement in any case. Placing an S^D at the start of conditioning was a practical device adopted after preliminary experimentation to insure that *S*s responded fre-

[1] This paper represents a modification of part of a Ph.D. dissertation, "Parameters in Human Operant Discrimination," submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the subject of Experimental Psychology, Harvard University, 1953. The writer expresses his sincere appreciation to Drs. W. S. Verplanck, B. F. Skinner, Floyd Ratliff, and C. B. Ferster whose direction and advice contributed much to the formulation of this paper.

[2] Now at Dartmouth College.

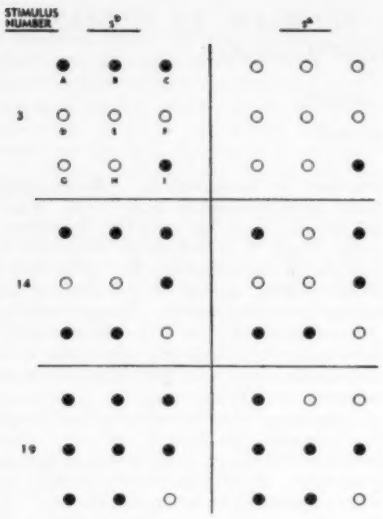

FIG. 1. Six representative discriminative stimulus patterns.

quently enough to meet the higher ratio schedules of reinforcement.

Recording.—Key tapping was recorded cumulatively on an electrical counter; the number of responses was transcribed on a protocol sheet each time stimulus cards were changed. Key tapping was also recorded separately on cumulative recorders while S^D's and S^{Δ}'s were presented. Two complete records of this response were so obtained. The cumulative records provided for quick comparisons and the observation of momentary changes in rate of response. The numerical record was convenient for mathematical treatment. Verbalizations were recorded verbatim.

After the second stimulus card had been shown to *S*, *E* asked, "Have you any ideas about what makes a card 'right'?" The reply was recorded, and after every second stimulus card throughout conditioning and extinction *E* asked, "Have you any new ideas?" These questions served as probes to encourage comments from *Ss* who were otherwise reluctant to state their hunches.

When *Ss* in the 3-sec. group started to verbalize, presentation of stimuli was halted until they had finished. Otherwise, the entire stimulus series might be run off while *S* was making one verbal response. Since *Ss* were not penalized in other groups for speaking, some such procedure had to be introduced to avoid penalizing *Ss* in this group.

Subjects.—The *Ss* were 45 volunteers from two undergraduate courses in psychology at Harvard University, and were randomly assigned to experimental groups. Their mean age was 20 yr., with a range from 18 to 24 yr.

Instructions.—The following was read to each *S* at the beginning of the experiment: "In this experiment we are investigating the formation of a discrimination. Through the opening in the panel you see before you, I am going to present a series of cards. Some of them are right, and some of them are wrong. There is a rule that makes them right or wrong. When you think one of the cards of the right group is being presented, I want you to press the key before you on the table. You will have to guess at first . . . perhaps for several of the cards . . . before you will be able to tell for certain that a right card is being shown to you. Try to find out the rule for right cards. You will know when a right card is being shown to you, because you will always get at least one point for pressing the key when one is before you. *Do your best to get as high a score as you can.* You will never get points for a wrong card. Remember, you will have to guess at first, perhaps for several cards, before you begin to get hunches as to what the rule is. Remember too, it is only by pressing the key when a right card is up that you can raise your score on the counter. Press the key when you think a right card is up."

The foregoing was read to all experimental *Ss*. For *Ss* in the 30-sec. and 60-sec. groups, the following was added: "You can sometimes get more than one point during the time that a right card is before you, so you must press the key repeatedly while a right card is up in order to get all the points you have coming to you."

The above was not read to *Ss* in the 3-sec. group, since they were allowed to make only one response per stimulus card, and instructions such as these would confuse them needlessly.

In addition to the instructions that were read to *Ss*, a summary of the instructions was typed on a card and this card was fastened to the panel for *Ss* to refer to at any time during the experiment. A statement was appended to the summary enjoining *Ss* to "Press the key only when you think a right card is up." This statement was added to imply punishment for responding in the presence of "wrong" cards. Without this instruction, some *Ss* in a pilot study continued responding through S^{Δ}'s although, when asked, they said that they knew the card was "wrong." Theoretical analysis of the necessity for punishment or threat of punishment in discrimination awaits additional experimental evidence.

Control groups.—Corresponding to the 1:1, 3-sec. group, and the 1:1 and 30:1, 60-sec. groups, three control groups were run to establish that *Ss* were actually under the discriminative

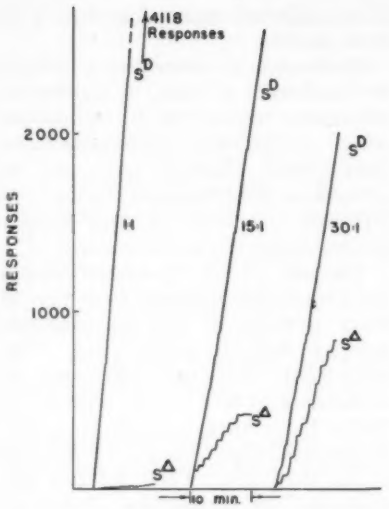

FIG. 2. Typical records in conditioning under fixed-ratio schedules with 60-sec. stimulus intervals.

control of the stimuli. Conditions for *S*s in these groups were identical to those in the corresponding experimental groups with the exception that there was no characteristic pattern distinguishing the S^D's from the S^Δ's. In the control groups, discrimination of the stimulus class was impossible.

RESULTS

Key-press response.—The records shown in Fig. 2 and 3, although those of *S*s in the 60-sec. 1:1, 15:1, and 30:1 groups, are representative of all *S*s run in this experiment. It will be noted that responding in S^D reaches a high rate early in conditioning. The lower the ratio of responses to reinforcement, the more rapidly the rate of responding increases, and this increase is reflected in the difference in total number of S^D responses under these three conditions.

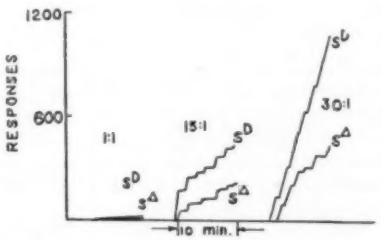

FIG. 3. Typical records in extinction under fixed-ratio schedules with 60-sec. stimulus intervals.

These records are also representative of curves generated by members of other species under fixed-ratio schedules of reinforcement. In S^Δ, bursts of responses at high rates occur that are followed by intervals of no responding. S^Δ curves in conditioning are characterized by progressively shorter bursts and progressively longer intervals between bursts. While these bursts become shorter and shorter, they never reached the actual ratio that was being used. In the 15:1 groups, *S*s generally responded 25 to 40 times in the presence of a particular S^Δ before they would stop. In the 30:1 group, they responded about 50 to 120 times before they would stop.

In extinction, the process reversed itself to some extent. The *S*s emitted more and more responses in S^Δ as time went by, if they responded at all. In other words, *S*'s discrimination of the ratio of responses to reinforcement progressively improved during conditioning, while it deteriorated during extinction.

Group means and variances of the numbers of responses under the various conditions were correlated, so a logarithmic transform was made of the data for purposes of statistical analysis. Conventional analysis of variance was precluded, but ordinal statistical techniques were applied where appropriate to check findings obtained with conventional small sample methods.

Neither the general pattern of response nor the subtle momentary changes described above are reflected

in the statistical analysis, but the effects of the controlling variables are nonetheless shown to be significant. The *Ss* discriminated between the two classes of discriminative stimuli as shown by the fact that they responded significantly more frequently while S^D's were presented than they did while S^Δ's were presented. The over-all probability that the *t* values of Table 1 would have occurred by chance is less than .01 in conditioning and less than .05 in extinction for the experimental groups.

If discrimination may be said to be more advanced with more responding in S^D than in S^Δ, then discrimination was farther advanced when lower ratios of responses to reinforcement had been used in conditioning. Longer presentations of discriminative stimuli also resulted in more advanced discrimination. χ^2 was used to determine whether different ratios produced significantly different over-all frequencies of responding to S^D's and S^Δ's in the three 60-sec. groups. The three groups were significantly different in both cases ($p < .001$). Differences in responding to S^D's and S^Δ's under continuous reinforcement in the three different interval groups were even greater.

Resistance to extinction is defined by the length of time, or number of responses, responding is maintained when no further reinforcement is forthcoming. Higher ratios of responses to reinforcement during conditioning resulted in significantly greater resistance to extinction.

The behavior of the control *Ss* was not appreciably different from that of those persons in the experimental groups during conditioning. The probability that the differences in responding to S^D's and S^Δ's shown in Table 1 for these groups in conditioning would have occurred by chance is less than .01. They did not discriminate in extinction, however. The probability that the differences obtained from these groups in extinction would have occurred by chance is near 1.00.

Verbalization.—Table 2 indicates the total number of verbalizations made in each group. It also indicates whether the verbalization occurred in S^D or S^Δ, in conditioning or extinction.

Different numbers of *Ss* in the vari-

TABLE 1

GROUP MEANS OF LOG R (KEY-TAPPING RESPONSE) AND TESTS OF SIGNIFICANCE

Interval	Ratio	Mean Log R—Conditioning						Mean Log R—Extinction					
		Experimental			Control			Experimental			Control		
		S^D	S^Δ	*t*	S^D	S^Δ	*t*	S^D	S^Δ	*t*	S^D	S^Δ	*t*
3 sec.	1:1	.95	.88	2.44*	.92	.90	3.85**	.84	.60	1.25	.88	.81	.50
30 sec.	1:1	3.12	1.06	14.86**				1.13	.91	.61			
	15:1	2.85	2.64	1.26				2.64	2.53	.88			
60 sec.	1:1	3.45	1.12	14.66**	3.41	1.31	10.33**	1.14	.91	2.11	1.39	1.42	.07
	15:1	3.39	2.81	4.27**				3.05	2.83	1.33			
	30:1	3.37	2.93	1.32	3.41	2.90	5.10**	3.14	2.90	3.35**	2.89	2.90	.00

* $p < .05$.
** $p < .01$.

TABLE 2

TOTAL AND TOTAL CORRECT VERBAL RESPONSES IN CONDITIONING AND EXTINCTION

Interval	Ratio	Responses in Conditioning								Responses in Extinction					
		Experimental				Control				Experimental			Control		
		S^D	S^Δ	Total	Total Correct	S^D	S^Δ	Total	Total Correct	S^D	S^Δ	Total	S^D	S^Δ	Total
3 sec.	1:1	3	6	9	2	3	3	6	0	5	1	6	1	0	1
30 sec.	1:1	5	7	12	3					8	7	15			
	15:1	5	4	9	2					4	4	8			
60 sec.	1:1	8	9	17	4	16	15	31	0	11	10	21	4	7	11
	15:1	10	8	18	4					9	6	15			
	30:1	6	9	15	3	8	10	18	0	12	6	18	10	5	15

ous groups verbalized correctly. The experimental groups were ranked in terms of the degree to which the key-tapping discrimination was advanced. They were also ranked in terms of the number of *S*s in the group who verbalized correctly. The rank-order correlation coefficient tau between these two measures was .89 ($p < .02$).

The three interval groups under the 1:1 ratio differed significantly in the total number of verbalizations emitted by *S*s, but significant differences were not produced by different ratios of reinforcement in the three 60-sec. groups.

Since each *S* can verbalize correctly only once during the experiment, the samples were too small for testing differences in number of correct verbalizations between groups.

DISCUSSION

Key tapping proved to be a sensitive variable in this study. As with similar responses studied in the rat and pigeon, responding was characterized by very high rates under fixed-ratio schedules of reinforcement, while extinction was characterized by rapid bursts of responding followed by intervals of no responding.

In the development of discrimination, responses in S^Δ were more resistant to extinction the higher the ratio of responses to reinforcement. Likewise, in extinction, all responding was more resistant to extinction where higher ratios had been used in conditioning. Thus it appears that human operant behavior is controlled in essentially the same ways by this type schedule of reinforcement as is the behavior of other species.

The *S*s discriminated more readily the longer the discriminative stimuli were presented to them. Not only did they have more time in which to accumulate large numbers of responses when responding was appropriate; but, the longer they had to respond, the more likely they were to state the correct conditions for discrimination.

The *S*s in the control groups discriminated *something* during conditioning. Since the stimulus patterns shown to them had no distinguishing characteristics, and since the apparent discrimination they did develop during conditioning was abolished at the onset of the extinction series, it seems clear that they were responding only to the presence or absence of reinforcement. That is, when these *S*s had responded several times after a new stimulus card was presented in conditioning and a reinforcement was given, they continued to respond for the remainder of the time that the particular stimulus was before them. If such a re-

inforcement was not forthcoming, they would cease responding in the presence of the particular stimulus. This undoubtedly accounts for the large differences in responding to the two stimulus classes in conditioning.

Presumably, *S*s in the experimental groups were controlled to an extent by the same kind of contingencies, but the fact that these *S*s maintained their discriminations in extinction demonstrates that they had been brought under the additional control of the discriminative stimuli. This observation suggests that studying the development of discrimination alone involves serious pitfalls that may be circumvented by also studying the course of extinction.

The number of *S*s who verbalized correctly in a group was directly related to the extent to which the key-tapping discrimination was developed. The more *S*'s key tapping was under the control of the discriminative stimuli, the more likely he was to state the controlling conditions correctly. The *S*s who did not verbalize correctly did not differ from those who did in the total number of times they pressed the key to S^D's and S^Δ's. It is clear that although the two measures covary, it is not necessary that *S* verbalize the conditions of discrimination in order to discriminate.

In brief, those conditions that produced rapid key-tapping discriminations were most likely to generate correct verbal statements of the conditions. Correct verbalization was not necessary for high resistance to extinction, since higher ratios produced greater resistance to extinction without generating many correct verbalizations.

Correct verbalizations represent "attainment of concepts" in the traditional concept formation experiment (**1, 2, 5, 7**). This experiment has demonstrated different ways of developing the verbalized concept, while using the same stimulus material, more or less efficiently. The traditional technique, as it was studied here, of presenting a discriminative stimulus or instance of the concept class, allowing *S* to respond once, and then reinforcing him appropriately, was the least effective. The sensitivity of the key-tapping response, permitting as it does the continuous observation of the process in time, is destroyed by abandoning the observation of changes in rate.

Summary

The *S*s in this experiment developed an operant stimulus discrimination of visual patterns of a level of complexity usually classed with concept formation. The extent to which *S*s discriminated was related inversely to the ratio of responses to reinforcement (1:1, 15:1, and 30:1) and directly to the length of time the discriminative stimuli were presented to them in conditioning (3, 30, and 60 sec.). The *S*s in control groups discriminated only the presence of reinforcement in conditioning.

Resistance to extinction was greater when higher ratios of responses to reinforcement and longer stimulus intervals were used in conditioning.

Verbalization was not a sensitive variable, but was related to the extent to which the key-tapping discrimination had advanced in conditioning. It was thus related to the parameters of the experiment in the same way as was the key-tapping discrimination.

References

1. Heidbreder, E. The attainment of concepts: I. Methodology and terminology. *J. gen. Psychol.*, 1946, **35**, 173–189.
2. Hull, C. L. Quantitative aspects of the evolution of concepts: an experimental study. *Psychol. Monogr.*, 1920, **28**, No. 1 (Whole No. 123).
3. Leeper, R. *Cognitive processes.* In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 730–757.
4. Skinner, B. F. *The behavior of organisms.* New York: D. Appleton-Century, 1938.
5. Smoke, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**, No. 4 (Whole No. 191).
6. Vinacke, W. E., The investigation of concept formation. *Psychol. Bull.*, 1951, **48**, 1–31.
7. Walk, R. D., The role of concept formation in discrimination learning. Unpublished doctor's dissertation, Harvard Univ., 1951.

(Received June 10, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

SHOCK FOR RIGHT AND WRONG RESPONSES DURING LEARNING AND EXTINCTION IN HUMAN SUBJECTS

CECIL M. FREEBURNE AND MARVIN SCHNEIDER

Bowling Green State University

It was found by Freeburne and Taylor (**2**) that animals shocked individually for both right and wrong responses learned a discrimination problem significantly faster than the same problem was learned by animals not shocked for either response. Their study arose from the observation that in the controversies concerning the effects of punishment in learning, investigators had provided shock for right responses only, or shock for wrong responses only, in such a way that the shock itself could serve as a cue to the correctness of the response, and so might come by way of differential secondary reinforcement to bring about the faster learning shown in the shock groups. Articles by Muenzinger (**8**), Wischner (**11**), and Spence (**10**) are relevant. The results obtained by Freeburne and Taylor were interpreted as support for Muenzinger's contention that shock anywhere after choice facilitates the learning of the response, inasmuch as they removed the possibility of differential secondary reinforcement by giving shock for both right and wrong responses to the same *Ss*.

The present study was designed to compare the effects upon learning of shock for right responses only, and shock for wrong responses only, as in the Muenzinger and Wischner studies, and of shock for both responses, as in the Freeburne-Taylor study, in an experiment using human *Ss*. It was thought probable that the results would be consistent with those previously obtained with rats under comparable conditions. However, even if this were the case, there would remain the question of the relative status of the concepts of facilitation and secondary reinforcement. Either concept could be called upon in explanation, except that facilitation might seem preferable to secondary reinforcement in the case of *Ss* shocked for both responses.

It happens that speed of adoption of a response is only one way of measuring learning. Slowness to give up making the response may also be used, i.e., extinction. It may well be that shock does "facilitate" learning, but that the relationship of the shock to some rewarding state of affairs is more important in maintaining the learned response. It seemed that carrying the present study into an extinction phase might shed some light on this question.

METHOD

Subjects.—Thirty-five male and 35 female volunteer undergraduate students from psychology classes at Bowling Green State University were individually assigned by name through use of a table of random numbers to one of seven groups with an *N* of ten each. Each group included five male and five female *Ss*. They were all assigned to group and condition before any were called into the laboratory.

Apparatus.—The *S* and *E* sat on opposite sides of a table with a barrier placed so that *E* was visible to *S* only from the shoulders up, while *E* could see *S* from tabletop up. In front of *S* was a blue-enameled box, 7 in. square and 3 in. high. On the box were two metal buttons attached to microswitch blades, which *S* could depress to indicate his choices of right or left, and a red light with which *E* could indicate to *S* that the next response was to be made. The switch blades actuated an oscillator, which sounded a tone of 512 cps through a 5-in. speaker

placed near the barrier on *S*'s side of the table, and they also activated a variable-intensity shock apparatus, which delivered shock by way of the metal switch blade and a metal rod held in *S*'s free hand. The function of the shock was that of punishment, whereas the tone served as the reinforcing agent. Both shock and tone were controlled from *E*'s side of the barrier, as was the order of correct and incorrect choices. When tone and shock were to be given to *S*, they followed automatically upon *S*'s choice, shock immediately and tone .5 sec. after. The shock was on as long as *S* held the key down. A sponge-rubber mat under the keys made it necessary for *S* to press so hard on the key that he could not avoid shock. The tone was of 1-sec. duration. Switches which had to be operated while *S* was present were noiseless, and could not be seen by *S*. Two lights on *E*'s box corresponded to *S*'s depressers, and indicated *S*'s choices to *E*.

Procedure.—In the learning phase of the experiment there were four conditions: NS, or no-shock control; SR, shock for correct responses; SW, shock for incorrect responses; and SB, shock for both correct and incorrect responses. One group of ten *S*s was assigned at random to the control condition, and two groups of ten (total = 20) to each of the three shock conditions, also at random. The *S*s were required to learn a temporal maze of 20 right-left choices. The series of choices was: RLRR LLRL LRLR LLRR LRRL. The learning criterion for this task was two successive series of 20 correct choices each.

In the extinction phase, one of the two groups in each shock learning condition was randomly assigned to receive no shock during extinction, while the other group in the same learning condition continued to receive shock during extinction just as they had received it during learning, and for the same responses. The extinction conditions were six: SR-SR; SR-NS; SW-SW; SW-NS; SB-SB; and SB-NS. In all extinction conditions, the previously reinforcing tone was omitted.

Upon being seated before the apparatus, *S* was told, if he was in a shock group, that part of the apparatus was shock apparatus. He was not told whether he would be shocked or when it might occur, even if he asked questions about the shock. In an attempt to hold down anxiety in the control group, control *S*s were told nothing of the shock apparatus. In shock conditions, before learning began, the current was adjusted for the individual *S* so as to produce a strong finger withdrawal. The *S* was asked to contact the shock a number of times so that the intensity of the shock could be increased until it was felt that the withdrawal was a function of the unpleasantness of the shock, as well as of anxiety. During *S*'s period in the laboratory, the strength of the shock was increased slowly but steadily until *S* complained of the increase, after which it was still increased, but more slowly. This was done to offset adaptation of *S* to the shock.

Instructions read to *S* included a statement of the task, including the information that the tone would sound only when a correct response was made. The shock was not mentioned in these instructions. It was explained that there was a series of 20 right-left choices to be learned, and that *E* would indicate to *S* each time that 20 choices had been made. The remainder of the instructions touched upon the mechanics of the performance. After *S* began his learning series, instructions were repeated if asked for, but no other questions were answered.

When the learning criterion was reached, the tone was cut off and the extinction begun by inserting a card into the tone-control relay on *E*'s control box. The *S* could not see or hear this done, nor was anything said to him about it. If *S* asked questions directly concerning the lack of tone, he was told to continue making choices, and that the tone would sound when he made a correct choice. This statement was repeated as often as it was specifically asked for, but no other instructions were given at this point. The session was concluded when *S*'s responses had dropped back to the chance level of 10 correct and 10 incorrect responses during a single run of the series of 20 choices. The reaching of this criterion was not taken to mean that *S* no longer "knew" the pattern. Many *S*s could recite the sequence even days later. However, the behavior of responding with the previously learned series of choices was failing to occur. While this is perhaps not the usual definition of extinction in such studies as the present one, it has been used before (9), and seems to fit the requirements of the present study.

It was considered likely that the SR-SR and the SW-SW groups might not extinguish. Both groups, having learned the series to a stringent criterion, might continue merely to make the learned responses even though there was no longer any tone to indicate correctness of responses. In the SW-SW group, this would mean continuing in extinction to make those responses which in learning had not only achieved the tone, but had also avoided the shock, which state of affairs continued in extinction for this group. In the SR-SR group, it was possible that the shock had acquired secondary reinforcing properties through association with the tone, and thus these *S*s might continue to make the shocked response even in the absence of the tone.

Therefore, the other five groups were run

before these two, and a prediction formula based on the *t* test of the significance of differences between means was used to compute a cutoff point. The largest mean and *SD* of the first five groups run were used for this purpose (the SW-NS group). On the assumption of equal group variances, if all members of the last two groups ran to this cutoff point without extinguishing, the differences between those groups and the others would be significant at the 1% level of confidence. Although the use of such a cutoff point for groups in which some members might extinguish before reaching that point might cause complications, it was deemed necessary to use such a cutoff point, since *S*s could not be run indefinitely. The cutoff point computed on the basis just outlined was 77; i.e., a given *S* was stopped after 77 extinction runs if he had not yet met the extinction criterion.

Results and Discussion

Before any *S*s were run in the present study, it was thought consistent with the general trend of pertinent literature (**1, 7**) to make the following predictions relative to the learning conditions: (*a*) the shock groups would all learn in significantly fewer trials than the no-shock group; (*b*) the shock-right and the shock-wrong groups would not differ significantly in numbers of trials taken to learn; and (*c*) the shock-both group would take significantly fewer trials to learn than would the other two shock groups. An *F* test of the learning data proved to be significant beyond the .001 level of confidence. Although several of the intergroup variance ratios were significant, this level probably indicates real differences among group means (**4**). However, the use of the *t* test for further analysis was questionable, and so the significance of the differences among groups was tested with the *U* test (two-tailed), as described by Mann and Whitney (**6**). The results of this analysis are presented in Table 1.

It may be seen that the predictions stated above were fulfilled in every comparison but one—Group SB did not learn significantly faster than Group SW, although it did learn faster. The conclusion seems justifiable, for this phase of the study, that the relationship between reward and punishment for human *S*s, under the conditions of the present study, closely resembles that previously observed in rats.

TABLE 1

Means, *SD*'s, and Results of *U* Tests for Learning Conditions

Group	Mean	*SD*	*U* Tests		
			SW	SR	SB
NS	48.70	34.29	151.5**	151.5**	177.0**
SW	20.40	9.60		196.5	238.0
SR	21.25	6.53			276.0*
SB	16.95	4.89			

* $p < .05$.
** $p < .01$.

In extinction, it appears that the continuance of a given response might depend not only upon the level of excitatory potential related to the response at the time the learning criterion is reached, but also upon the continuance of differential secondary reinforcement during extinction (**1**). In situations involving punishment, it has been proposed (**7**) that the punishment produces anxiety, which heightens the general drive state of the organism. Hull (**3**) would then suggest that this in turn leads to a greater level of excitatory potential when combined with a given level of habit strength, and thus to earlier consistent performance of the observed response. The rewarding state of affairs is presumed to be anxiety reduction occurring upon cessation of punishment. It may well be that Muenzinger's concept of facilitation corresponds to this. However, if onset and cessation of punishment occur within time intervals known to be effective in conditioning, punishment itself may come to serve as secondary reinforcement, and responses may become conditioned to shock acceptance as well as to shock

avoidance. Especially may this be true with human *S*s, who know that the shock will be of limited severity. The reward, the anxiety reduction, may come from finding the shock within tolerance limits, with *S* left intact and still functioning. If such punishment has been consistently associated with given responses, then it might serve to maintain those responses in the absence of other specific agents which reinforced them during learning. Groups extinguished without shock, then, as under the conditions of the present study, might be expected to extinguish more readily than those extinguished with shock because there would be no inhibiting carry-over from the previous learning situation (**5, 9**). Following this reasoning, and prior to running any *S*s, it was predicted that in those conditions in which shock had accompanied learning in such a way that it could serve as a cue to correctness of response, or could acquire secondary reinforcing properties, extinction under the same shock conditions would require significantly greater numbers of trials than would be required where the response was extinguished without shock, or where shock continued to accompany both responses.

An *F* test of the data for the extinction conditions proved to be significant beyond the .001 level of confidence. Again, the *U* test (two-tailed) was used to test group differences, for reasons stated above. The means, *SD*'s, and results of *U* tests for the extinction conditions are given in Table 2.

All but two of the comparisons pertinent to the prediction just stated concerning the extinction conditions in the present study bear out that prediction. In one of the exceptions, the comparison of Group SW-NS with Group SR-SR, the difference is in the expected direction. In the other exception, the comparison of Group SB-SB with Group SR-SR, the difference is in the expected direction and is significant beyond the .10 level of confidence.

In extinction, *S* can be seen to be operating under a complex motivational state. He had volunteered to serve as *S*, and so was presumably motivated to do whatever *E* told him to do in the experimental situation. This is evidenced by the fact that *S*s persisted until told by *E* that they were through, even though some *S*s required hours for the achievement of the two criteria set up for learning and extinction. The *S*'s responses in learning were influenced by the noxious stimulus of the shock, and he can be assumed to have been motivated toward shock avoidance, since shock was made strong enough that *S*s complained of its strength now and then. Shock was intentionally increased in strength gradually to insure that this would continue to be so. The *S*'s responses in learning were also influenced by the occurrence or nonoccurrence of the tone, which

TABLE 2

MEANS, *SD*'S, AND RESULTS OF *U* TESTS FOR EXTINCTION CONDITIONS

Group	Mean	*SD*	*U* Tests				
			SW-SW	SR-NS	SB-NS	SB-SB	SR-SR
SW-NS	36.60	29.40	10.00**	73.50	75.00	72.00	44.00
SW-SW	77.00	†		00.00**	00.00**	00.00**	85.00*
SR-NS	8.60	7.26			61.50	71.50	16.00*
SB-NS	8.50	7.38				48.00	13.00*
SB-SB	14.80	20.76					76.00
SR-SR	41.80	26.82					

† All *S*s in this group ran to 77-trial cutoff point without extinguishing.
* $p < .02$.
** $p < .01$.

served as the reinforcing agent during the learning series.

Since *S* was told that the problem in extinction was the same as it was in learning, we can assume that the abovementioned factors were still effective in his motivational state. In addition, *S* soon learned that no response achieved the tone, and *S*'s verbalizations left no doubt that he realized he had a new problem to solve. The attempt to solve this new problem led toward extinction of the response established in original learning. The solution of this new problem was complicated by the fact that the shock and the tone had come to stand in certain relationships with one another, depending upon the particular learning condition for a given *S*. For some *S*s, the shock had served as a cue to the correctness or incorrectness of their responses during learning. In the SW learning conditions, the shock-avoidance tendency was reinforced, because in making the response reinforced by the tone, *S* avoided the shock. In the SR learning conditions, any shock-avoidance tendency had to be overcome to achieve successful solution, and the shock could have acquired secondary reinforcing properties, as well as serving as a cue. In the SB conditions, likewise, the shock-avoidance tendency had to be overcome for solution to be reached, but there was no way to avoid the shock, and the shock could not be differentially associated with either right or wrong responses. However, anxiety and rewards associated with its reduction might be assumed to speed the solution of the problem. The data in Table 2 appear to support the suggestion that these considerations carry over into extinction.

To conclude, we may say that the results of the learning phase of the present study are consistent with the literature touching upon similar problems, and that with these human *S*s, the shock groups did learn faster than the no-shock groups. While it is true that other explanations might be made, the present writers wish to rest their explanation upon the concept of differential secondary reinforcement, where conditions make this possible, as in the SR and SW groups. Where this is not possible, as in the SB group, Muenzinger's concept of general facilitation, or the concept mentioned by Hull, Spence, and Miller of drive heightened by way of anxiety may be called into the explanation. Since one would predict the same outcomes using either concept in the latter case, there seems to be little difference between the two. The same general considerations seem to apply to the extinction data. Actually, choice of concepts does not appear to be critical in the present study. It is undoubtedly true that the human *S* employs more complex processes in problem solution than does the rat. Nevertheless, the predictions made relative to the conditions of the present study were made by way of reasoning through the use of the concepts indicated, which were founded upon the results of studies using rat *S*s. The high proportion of successful prediction of human behavior here encountered suggests a communality of behavior determinants which should be heartening to behavior theorists.

Summary

The present study investigated the effect of shock for right and wrong responses with 70 human *S*s, in both learning and extinction. The *S*s learned a pattern of 20 right-left choices in a temporal maze under four conditions: no-shock, shock for right responses, shock for wrong responses, and shock for both responses. Correct responses were reinforced by a tone of 512 cps, which was omitted during extinction.

Learning under all three shock conditions was significantly faster than with no shock, and the shock-both group learned significantly faster than the shock-right group. The shock-right and shock-wrong groups did not differ significantly. These results are consistent with previously published studies using animal *S*s.

In extinction, the shock-learning groups were each divided into two subgroups, one extinguished without shock and one extinguished with shock continuing as it had in learning. The results here show the effects of punishment to be relatively complex. Where the conditions for differential secondary reinforcement occur, so that the shock may serve either as cue or agent for such reinforcement, extinction tends to be significantly delayed over nonshock

extinction conditions. The fastest extinction was shown by the group shocked for both responses in learning, and then shocked for both responses in extinction.

References

1. Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 568–612.
2. Freeburne, C. M., & Taylor, J. E. Discrimination learning with shock for right and wrong responses in the same subjects. *J. comp. physiol. Psychol.*, 1952, **45**, 265–268.
3. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
4. Lindquist, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
5. Maatsch, J. L. Reinforcement and extinction phenomena. *Psychol. Rev.*, 1954, **61**, 111–118.
6. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
7. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435–469.
8. Muenzinger, K. F. Discussion: concerning the effect of shock for right responses in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 201–203.
9. Ni, C. An experimental study of the influence of punishment for errors during learning upon retention. *J. comp. Psychol.*, 1934, **17**, 279–301.
10. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54**, 1–8.
11. Wischner, G. J. The effect of punishment on discrimination learning in a non-correction situation. *J. exp. Psychol.*, 1947, **37**, 271–284.

(Received May 27, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

DISCRIMINATIVE SKILL AND DISCRIMINATIVE MATCHING IN PERCEPTUAL RECOGNITION[1]

JEROME S. BRUNER
Harvard University

GEORGE A. MILLER
Massachusetts Institute of Technology

AND CLAIRE ZIMMERMAN
Wellesley College

An important aspect of any recognition or identification task is one's knowledge of the set of alternative items from which the stimuli are drawn. In general, we can respond most accurately when we have the fullest knowledge of what stimulus events are possible. In many recognition experiments, however, the set of possible stimuli is left ambiguous at first and only after considerable experience in the situation does one become familiar with the range and variety of stimuli that are used and master a set of responses appropriate to them. In such experimental situations there is usually a considerable improvement in performance as one narrows down the initially ambiguous range of possible stimuli to the particular set that is actually presented.

At the same time, one's skill as such in recognizing items is improved by practice. Learning to recognize briefly or peripherally exposed materials, or learning to recognize items masked by noise or glare—these are skills that also improve with practice regardless of the items to be recognized. In the present experiment, improvement in knowledge of items to be recognized is measured as a function of *S*'s mastery of the list of stimulus items. Thus two measures were taken, one of the degree to which the stimulus alternatives were known and a second of the accuracy with which these same stimuli could be recognized.

[1] This research was supported jointly by the Laboratory of Social Relations and the Psycho-Acoustics Laboratory of Harvard University under its ONR Contract N5-ori-76 (Project NR 142-201, PNR-172).

Method

The procedure utilized has as its objective the testing of two psychological processes: (*a*) the ability to *recognize* a list of words presented one at a time in a noisy channel (at a signal-to-noise ratio of −18 db) and (*b*) the ability to *recall* the same words. Four groups were used, varying in size from 13 to 15 *S*s. All groups received identical treatment, to be described presently, save in one respect: the length of the word list they were required to recognize and recall varied. The lists contained either 8, 16, 32, or 64 words, all monosyllables drawn from the Egan list (2) and phonetically counterbalanced to reflect the average frequencies of occurrence of the various speech sounds in ordinary discourse. Each of the groups receiving lists of a certain length were subdivided in such a way that all 64 different words were presented. That is to say, in the group given a list of 8 words, for example, 8 subgroups were tested, each given a different portion of the 64-word list. Thus, the total scores of the 8-word group are based on the same stimulus materials as were presented to the 64-word group. The same procedure was used with the groups given 16-word and 32-word lists. Thus, four subgroups of the 16-word group had to be tested, and two subgroups of the 32-word group.

The sequence of recognition and recall tests was as follows: (*a*) *Recognition Test I:* Words were presented in the presence of a masking noise and *S* wrote down any words that he heard. The *S* was given no indication of the nature of the words. (*b*) *Recall 1:* Words were presented without the masking noise. The *S* was told that his task was to recall these words immediately after the entire list was read. At the end of the presentation *S* wrote all the words he could recall. (*c*) *Recognition Test II:* Recognition test repeated as in I. (*d*) *Recall 2:* Presentation of list and recall test repeated as in Recall 1.

(*e*) *Recognition Test III.* (*f*) *Recalls 3–4.* (*g*) *Recognition Test IV.* (*h*) *Recalls 5–8.* (*i*) *Recognition Test V.* (*j*) *Recalls 9–16.* (*k*) *Recognition Test VI.* (*l*) *Recalls 17–32.* (*m*) *Recognition Tests VII–X.* (*n*) *Recognition Tests XI–XIV:* The *S* had a written list of the test words in front of him to refer to. Thus, *S* received a total of 32 recall tests and 14 recognition tests. One of the recognition tests was taken with no knowledge of the words contained in the list, 10 with varying degrees of knowledge of the list, and 4 with the list present so *S* could check his response. The order of presentation of words in both kinds of tests was random and varied from one test to the next.

The recognition tests were conducted as follows. The list presented consisted of the 8, 16, 32, or 64 words. In each instance, a quarter of the words, selected at random, were repeated once in order to prevent *S* from operating by a process of elimination. The *S* was told: "We are interested in finding out how well some common words can be heard under conditions where there is a great deal of noise. By noise I mean a sound like this (masking noise demonstrated). I am going to read you a list of 8 (16, 32, or 64) words with the noise turned up. After I read *each* one, write down the word you think it is. I'll introduce each word with the phrase, 'The word is. . . .' Eight (16, 32, or 64) and only eight different words will be read; some of them may be repeated. Turn to page 1 of the booklet. Any questions?"

The *S*s were given from 4 to 6 sec. between word presentations for writing their responses. On the last four recognition tests, as mentioned earlier, *S*s were given the lists in written form and told they could refer to these lists. The *S*s received the words in noise through earphones (PDR-8). The speaker, a woman, monitored the signal-to-noise ratio by a VU meter. The voice channel had a frequency response essentially uniform from 200 to 6500 cycles. The masking noise had a uniform power spectrum through this range of frequencies and the over-all signal-to-noise ratio was set at −18 db.

In the recall procedure, the words were read over the same channel without noise present. The *S*s were told that these were the same words that were being presented in the articulation test. The words were presented at the rate of one per second. Upon completion of the reading of the list, *S*s were given an amount of time determined by the length of the list, during which they were instructed to write down as many words as they could remember. The length of time permitted was set at 3 sec. per word, e.g., the 8-word group was given 24 sec. They were told that it was not necessary to recall words in the same order in which they had been read.

The *S*s of the 8-word, 16-word, and 32-word groups completed the experiment in one sitting requiring .5 to 2 hr. The 64-word group was given a 10-min. break after Recognition Test VI, the total time required being approximately three and a half hours.

Results

As expected, the number of words recalled on successive trials increased—in sum, learning occurred. The average number of words recalled on successive trials, expressed in percentage of recall, are presented graphically in Fig. 1A. Again, as expected, the shorter lists were more easily learned and a larger percentage of the words in them was recalled.

The four learning curves are of the conventional form, showing a rapid initial increase and then leveling off. After 32 trials the four groups reached

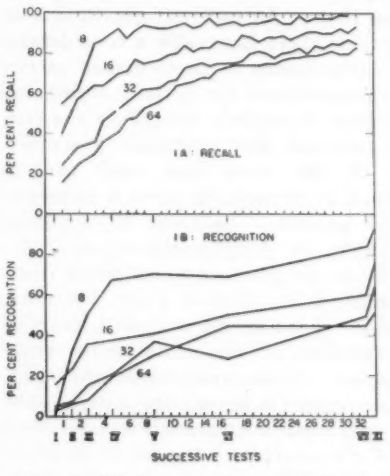

Fig. 1. (A) Percentage of recall for four different lengths of list is plotted as a function of the number of recall tests. (B) Percentage of recognition for four different lengths of list is plotted as a function of the number of recognition tests. The abscissa represents the order in which successive recall and recognition tests were presented.

levels of approximately 99% for the 8-word list, 94% for the 16-word list, and 85% and 82% for the 32-word and 64-word lists. It seems likely that under the conditions of this experiment the asymptotes of the curves for the 32-word and 64-word lists would fall below 100% recall; these longer lists required longer experimental sessions with attendant boredom for Ss.

An interesting relation emerges between the lengths of the lists and the number of trials required to reach a given level of recall. For example, 50% mastery was exceeded on Trial 1 by the 8-word group, on Trial 2 by the 16-word group, on Trial 5 by the 32-word group, and on Trial 8 by the 64-word group. That is to say, the number of trials needed to reach criterion was a constant fraction (approximately $\frac{1}{8}$) of the length of the list. At the 75% recall level this fraction is of the order of .4, although more variability is evident—the 64-word group seems to reach the higher criterion in fewer trials than this simple hypothesis would predict. The individual recall functions have

TABLE 1

CORRECT RESPONSES AND ERRORS IN RECOGNITION TESTS ON FOUR LENGTHS OF LIST*

Length of List	Recognition Tests							
	I	II	III	IV	V	VI	VII–X	XI–XIV
Correct Responses								
8	0.2	3.3	5.1	7.0	7.2	7.0	8.3	9.2
16	3.5	4.5	7.2	7.4	8.5	10.1	11.8	14.8
32	2.3	3.7	7.1	8.6	14.6	12.0	20.0	25.5
64	2.3	5.3	6.8	17.5	25.1	36.1	36.1	41.0
List-Word Errors								
8	0.0	0.5	0.5	0.6	0.7	0.9	0.8	0.5
16	0.1	0.8	1.9	4.2	3.9	2.9	3.2	4.0
32	0.8	3.5	4.6	5.7	5.3	9.8	6.4	8.2
64	2.0	5.5	6.0	8.0	10.2	11.5	9.1	11.8
Nonlist-Word Errors								
8	3.7	3.6	2.1	0.9	0.5	0.4	0.1	0.0
16	10.8	6.8	5.1	3.7	2.6	1.4	0.8	0.1
32	24.3	21.0	18.0	14.2	9.4	5.1	2.4	0.8
64	46.1	36.4	34.4	26.5	17.3	9.1	12.3	1.6
Blank Responses								
8	6.1	2.6	2.3	1.5	1.6	1.7	0.8	0.3
16	5.6	7.9	5.8	4.7	5.0	5.6	4.2	1.1
32	12.6	11.8	10.3	11.5	10.7	13.1	11.2	5.5
64	29.6	32.8	32.8	27.9	27.4	23.3	22.5	25.6

* In order to convert these frequencies into percentages, the frequencies for the 8-word list should be divided by 10, those for the 16-word list by 20, those for the 32-word list by 40, and those for the 64-word list by 80; then the quotients should be multiplied by 100.

been discussed by Miller and McGill (4) and will not concern us here.

As for recognition scores, they show a marked improvement with practice. For the 8-word list, for example, the recognition scores increase from 2% to 92%; for the 64-word list the improvement is from 3% to 51%. Table 1 contains a summary for the lists of various length of the average number of words correctly recognized and of the frequencies of the different kinds of errors that occurred. Data for recognition on Trials VII–X are grouped to obtain a better estimate of asymptotic performance level when *Ss* operated without word lists before them. Similarly, the data for Trials XI–XIV are grouped, this time better to estimate optimum performance after learning has leveled off and when *Ss* operate with a word list before them as an aid. Figure 1B presents data on the percentage of the list correctly recognized, these data being plotted on a common abscissa with the percentage of recall scores.

What of the relation between recognition and recall? Clearly the two are positively correlated. Consider an *S* operating in the recognition procedure before he has obtained any information about the list save that it is composed of English monosyllabic words. Under the conditions of this experiment, one would expect no difference as a function of length of list. Consistent differences in recognition begin to emerge only after *S* has studied the list. When *S* has achieved mastery, recognition performance is consistently better for the shorter lists. Although the same words appear in the short as in the long lists, it is nonetheless evident that discrimination is easier when *S* is set to select 1 out of 8 rather than 1 out of 64 alternative words (3).

The relation between recall and recognition is made explicit in Fig. 2. In Fig. 1, both recall and recognition are plotted as a function of number of trials. It is therefore possible to relate level of recall directly to level of recognition independently of the number of trials required to reach these levels. It is such direct functions that have been plotted in Fig. 2. In order to construct those functions, the curves in Fig. 1 had first to be smoothed, an operation performed by visual inspection. The smoothed functions were then related directly to each other, e.g., for the 32-word list, a 40% recognition score was obtained on the same trial that a 75% recall score was obtained, whereas for the the 8-word list, the 40% recognition score was obtained when the recall score reached 68%.

FIG. 2. Smoothed functions relating percentage recall to percentage of recognition for four lengths of lists.

For recall scores below 60%, a single function can describe the relation between recognition and recall for all lengths of list; over this range, the recognition score for the noise level employed is approximately half of the recall score. Superior recognition of words in the shorter lists becomes evident only after *S* is able to recall more than half of the list.

In the recognition scores considered thus far three kinds of mistakes have been thrown together: (*a*) responses in which the presented word is identified as a word that is not on the list, (*b*) responses in which the presented word is mistaken for another word that is on the list, and (*c*) a "blank" response to the presented word. One would expect that the relative frequencies of the three kinds of error would change as learning progresses. In Table 2 (B, C, and D) the three kinds of recognition errors are presented for the lists of different length. Note first that there is a steady decline of nonlist errors for all lengths of list. List errors, on the other hand, tend to increase even though the total number of errors being made is declining. The number of blank responses remains approximately constant for all lists, save the shortest. In the 8-word case, there is a decline in such blank responses, possibly explainable as adaptation to the articulation testing situation in which this group had considerably less experience than other groups, since for them each recognition trial consisted of but ten words.

The marked improvement in recognition scores can, then, be accounted for by the steady elimination of extraneous or nonlist responses as learning progressed.

Discussion

The *S*s who were tested with the 64-word list received eight times as much practice in recognizing words in noise as did those tested with the 8-word lists, four times as much as the 16-word *S*s, and twice as much as the 32-word *S*s. This learning process—learning to distinguish the spoken sounds from the interfering sound—must be contrasted with learning the words on the respective lists. As a consequence of the experimental design, the 8-word group had the best opportunity for mastering the recall task and the poorest opportunity to master the technique of word recognition. How critical this fact is for the interpretation of the results will shortly become evident.

What relation would one expect between recognition and recall in this situation? Two alternative hypotheses present themselves. According to the first of these, recognition and recall are unrelated; recognition improves with practice in recognizing and is unaffected by the extent to which the person can recall the items to be recognized. If this were the case, one would expect that those who worked with 64-word lists would show more improvement in recognition because they had more practice at it.

According to the second hypothesis, recognition is a process of matching the stimulus input to one of a set of recalled (i.e., available) items. This matching process gets more difficult as the number of items increases. For example, a listener who recognizes that a stimulus might be any one of 8 items on a 64-word list is in some sense performing at the same level as one who correctly recognizes that the stimulus is a particular one of the items on an 8-word list. According to this hypothesis, one would expect that *S*s who worked with 8-word lists would show more improvement in recognition because at any point in learning the same discriminative capacity would produce a higher percentage of correct matchings.

The two hypotheses, therefore, lead in opposite directions. The first predicts greater improvement for the long lists, whereas the second predicts greater improvement for the short lists.

If we look only at the point where learning has reached a maximum and *S*s have the complete lists before them, we see that the second hypothesis provides an adequate description of the data. In terms of information theory, a stimulus from the 8-word list represents 3 bits of information, one from the 16-word list represents 4 bits, from the 32-word and 64-word lists, 5 and 6 bits, respectively.

That is to say, 3, 4, 5, and 6 binary decisions are needed to match a particular stimulus input correctly when there are 8, 16, 32, and 64 alternatives. At maximum performance *S*s were able to get 93% of the 3-bit, 75% of the 4-bit, 64% of the 5-bit, and 51% of the 6-bit discriminations. If we take these proportions as a rough approximation, therefore, we find that *S*s were able to handle about 3 bits of information in every case. In sum, the second hypothesis is consistent with the asymptotic performance.

At preasymptotic levels of learning, however, neither hypothesis alone is adequate. Indeed, when recall scores are less than 60%, recognition scores in all four groups are indistinguishable. How to explain this result?

Note that the two hypotheses are not mutually exclusive. Nothing precludes the possibility that both types of learning go on simultaneously. If both types occurred, one might well expect that the advantages accruing to the longer lists from learning to recognize would counterbalance the margin to be gained by the ease of learning to recall the shorter lists. Thus the curves in Fig. 2 probably represent the combined effects of both forms of learning. If equal amounts of discrimination practice had been given to all groups, we would expect that the shorter lists would have been recognized more accurately at all stages of learning.

Finally, we must raise anew the question of what is involved when identification or recognition improves with practice. At least two aspects of recognition must be distinguished. One of them can be described as *discriminative skill*. The acquisition of discriminative skill involves such things as abstracting relevant from irrelevant stimulus dimensions, learning when to listen or where to look, and in most general terms learning to overcome internal and external distractions. Relatively little is understood about learning such discriminative skill, but it may well be that such classical experiments as the abstraction studies (e.g., 1) can shed light on the phenomena involved. A second aspect of recognition is *discriminative matching*, a process of sorting the stimulus inputs into appropriate categories. Matching improves when a person learns what categories to use. The development of discriminative skill makes it possible to match inputs to categories once the categories have been established. In sum, two learning processes may develop simultaneously and complement each other in the recognition task. It is only after these two learning processes have stabilized that performance in recognition can be serviceably described in terms of the concepts of information theory.

Summary

Lists of 8, 16, 32, or 64 monosyllabic words were learned by a free recall method. Before, during, and after this learning period, tests were made of *S*'s ability to recognize these words spoken in the presence of a severe masking noise. Both recall and recognition scores increased with practice. The increases were largest and most rapid for the short lists. The improvement in accuracy of recognition is attributed to two different learning processes: (*a*) the acquisition of discriminative skill, which seems to involve an adjustment by *S* to relevant aspects of the recognition task, and (*b*) the improvement in discriminative matching, which results from knowledge of the set of alternative possibilities to which the individual stimulus events must be matched.

References

1. Chapman, D. W. Relative effects of determinate and indeterminate *Aufgaben*. *Amer. J. Psychol.*, 1932, **44**, 163–174.
2. Egan, J. P. Articulation testing methods. *Laryngoscope*, 1948, **58**, 955–991.
3. Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. *J. exp. Psychol.*, 1941, **41**, 329–335.
4. Miller, G. A., & McGill, W. J. A statistical description of verbal learning. *Psychometrika*, 1952, **17**, 369–396.

(Received May 25, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

EXTINCTION AS DUE TO THE CHANGED DIRECTION OF A PSYCHOLOGICAL FORCE

KOH JOHDAI

Hiroshima University

Suppose that S is in a "detour problem" which can be solved only by using the detour (D), i.e., a situation where a barrier (B) is placed between an incentive (G) and S. When S tries vainly to make a dash at the barrier from his present position (A), we can make the following statement about the direction of a psychological force $f_{A,G}$ in the field:

$$d_{A,B} = d_{A,G} \quad \text{and} \quad d_{A,D} = d_{A,-G}. \quad (1)$$

The S is unable, of course, to solve the problem as long as the force $f_{A,G}$ does not change its direction. But, when the correct solution has been found after some period of trial-and-error, then we can make another statement about the direction of the force at that moment (**7**, pp. 73, 253):

$$d_{A,B} \neq d_{A,G} \quad \text{and} \quad d_{A,D} = d_{A,G}. \quad (2)$$

In other words, we must deal, in such a case, with the distribution of forces at two moments (one concerning the situation before the solution and the other concerning the situation after the solution) between which a "historical process" is interpolated. This historical process involves, then, a kind of "extinction" of the dash response[1] and an occurrence of another response which replaced it. Extinction in this case can be viewed, therefore, as due to the changed direction of a psychological force $f_{A,G}$. In other words, the greater the changeability in direction of the force, the smaller the resistance to extinction, other factors being equal.

That extinction may be due to the changed direction of a force $f_{A,G}$ has not been examined, because extinction procedures used in the experiments up to this time have dealt merely with the conditioned response to be extinguished. If we provide a situation where an alternative response signifies extinction of the original conditioned response (a situation which is similar to the detour problem in some respects), we may be able to relate extinction to the changed direction of a psychological force. Even in a situation where such a substitute response has no room to take place, extinction of "locomotion" along the "path" which does not lead to the goal any longer may be viewed as due to the changed direction of a force (**6**, p. 148). For example, extinction after withholding of food (G) of lever-pressing response (L) of a rat (A) in a Skinner box can be thought of as due to the change of a force $f_{A,G}$ in direction from

$$d_{A,G} = d_{A,L} \quad (3)$$

to

$$d_{A,G} \neq d_{A,L}. \quad (4)$$

But, because in this case there exists no region X such that $d_{A,X} = d_{A,G}$ ($w_{A,X} \subset w_{A,G}$), change of the force $f_{A,G}$ in direction from $d_{A,L}$ to $d_{A,X}$ cannot be brought to light.

Only where extinction of a response is indicated by occurrence of another response can the above change of direction be directly observed. If

[1] Especially when the response can be viewed as due to the generalization of some past experience.

we can show the usual fact of less resistance to extinction after continuous reinforcement than after partial reinforcement (**4**) in this new kind of alternate route situation, then we shall have proved our thesis.

Method

Apparatus.—Serial responses in a wooden stylus maze (Fig. 1) during acquisition and extinction allowed measurement of response strength by the number of errors. The maze pattern was like Sartain's stylus maze (**9**) in all respects except that the present setup had small, .5 in. long metallic plates instead of stops

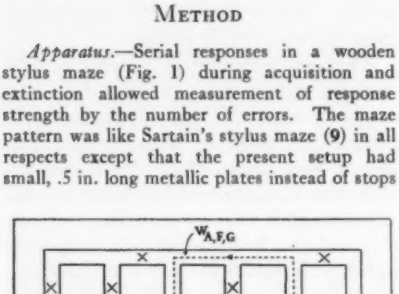

Fig. 1. Pattern of the stylus maze used. *A*, starting point at one time and terminal point at another time. The starting was indicated by an ordinary electric light, and arriving at the goal (*G*) was indicated by a red light which meant scoring 20 points. The experimenter could switch on at will either of these two lights whenever the stylus was placed on a point *A*. He also could prevent them from being lighted even when the stylus was placed on a point *A*. Whenever the stylus was brought to a point *E*, a green light was exposed at the window; whenever the stylus was brought to a point *F*, an orange light was exposed. *P* and *Q* indicate one-way doors which permitted locomotions only in the direction of arrows. Whenever the stylus was brought to a point *X*, *S* was given a buzzer shock. S indicates location of *S*; $w_{A,E,G}$, green path; and $w_{A,F,G}$, orange path.

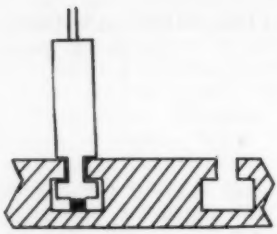

Fig. 2. Cross section through the wooden stylus maze (about two-sevenths actual size).

on the floor of the grooved pathway. The stylus on any of the plates closed the circuit of an electric light or a buzzer (see Fig. 2).

A black cloth screen was set up between *S* and the maze in such a way that he could move the stylus freely with his right hand but could not see the maze directly so long as he did not bend forward and peep. At approximately eye level of the screen a frosted-glass window 2.5 in. in diameter was lighted by differently colored lights when the stylus was brought to different points (see the explanation of Fig. 1). A head holder was used to keep *S*'s face directly before the window, and a buzzer was tied to the stock of the holder.

Subjects.—Thirty students[2] of Hiroshima University were used as *S*s. The average age was 21 with a range from 19 to 33 yr. They were divided randomly into three groups of equal number: one continuously reinforced group (RR), and two partially reinforced groups (RU and UR) different in patterns of partial reinforcement (**1**). The experimental procedure was the same for all *S*s except for patterning of reinforcement.

Procedure.—The experimenter and *S* sat opposite each other with the maze and an opaque screen between them. Each *S* was tested individually. The following instructions were read to each *S* at the beginning of the experiment: "Your task is to make a circuit of paths in a stylus maze from a starting point to the same place. You cannot see the maze during the experiment. Keep your head on the head holder and be careful to notice differently colored lights exposed at the window.

"Two independent paths are provided for you: 'green path,' the first step of which is characterized by starting to the right and then

[2] The total number of *S*s was 40. The results of 10 could not be used, however, because they misunderstood the instructions or took the orange path during the training series prior to the regular extinction procedure (see Table 1).

exposing a green light, and 'orange path,' the first step of which is characterized by starting to the left and then exposing an orange light. These two paths have no common parts except starting and terminal points. Whenever you turn aside from the true path, you will be given a buzzer shock. Then, you have to return the stylus to the true path without passing over the buzzer point and trace on to the terminal point.

"Your marks obtained in this experiment will be determined in the following way: You score 20 points if you have been able to make a circuit and thus light a red lamp (the experimenter exposed the red light at the window) when you have arrived at the terminal point. However, each buzzer (i.e., error) means subtracting one point. The higher score you make in repeated trials during the experiment, the better your results will be. So, try to make as good a score as possible. You need not trace fast because the time required for you to make a circuit has no bearing upon your score.

"Now, as preliminary training you have an opportunity to try each path once. This light (the experimenter exposed the ordinary electric light) means 'start' every time. First you may take the orange path. Hold the stylus with your right hand. Start! (At the same time the experimenter exposed the ordinary light as a signal of starting and then recorded the number of errors made during the trial by means of a counting apparatus, and exposed a red light when the stylus had been brought to the terminal point.) Now, you may try, this time, the green path (*S* made one trial in quite a similar way). We are ready to begin the main experiment where you can score. You may use *either* path in order to make as good a score as possible during the experiment.

"Briefly again: (*a*) a red light means scoring 20 points; (*b*) each buzzer means subtracting one point; (*c*) 20 points is the highest possible score in each trial; (*d*) there are provided two (orange and green) independent paths; (*e*) try to make as good a score as you can. In order to do so, you may use either path quite freely during the experiment."

After the above instructions, experiments were carried out under the three different conditions shown in Table 1. Letter R in Table 1 means reinforcement by exposing a red light, i.e., scoring 20 points, when the stylus was brought back to the point A. Letter U means nonreinforcement by exposing no light. One-minute rests were interpolated after the Series 1, 2, and 3.

Immediately after each trial, whether reinforced or not, *S* was informed of his score by the experimenter: 20 minus the number of errors in case of the red light (R) and 0 minus the number of errors in case of no light. Immediately after this the starting light was exposed, and *S* began a new trial. Each *S* in our experiment, therefore, just like a rat in a Skinner box, was placed at the starting point whenever he had finished a response.

The resistance to extinction of a green-path response was measured by the number of nonreinforced trials prior to the occurrence of an orange-path response.

TABLE 1

PROCEDURE OF TRAINING AND EXTINCTION FOR CONTINUOUSLY OR PARTIALLY REINFORCED GROUPS

Experimental Group			Terminal Training and Extinction (All Groups)
RR	RU	UR	
Series 1			R
			U_1
			U_2
			U_3
			U_4
R	R	R	.
R	R	R	.
R	R	R	.
R	R	U	.
R	U	U	.
R	U	R	.
			.
			U_n
Series 2			
R	R	R	
R	R	U	
R	U	U	
R	U	R	
Series 3			
R	R	R	
R	R	R	
R	R	U	
R	U	U	
R	U	R	

RESULTS AND DISCUSSION

Results of the preliminary training.—That the green path with 15 buzzer points was much easier than the orange path with 31 (see the *X*'s in Fig. 1) is shown by Table 2. It was, therefore, quite natural that all *S*s should use the green path consistently during the training period in order to make their scores and that it

TABLE 2
MEAN NUMBER OF ERRORS IN PRELIMINARY TRAINING

Path	N	Mean	SD
Green path ($w_{A,E,G}$)	30	8.70	3.33
Orange path ($w_{A,F,G}$)	30	22.66	5.88

was this response which was extinguished.

Resistance to extinction after continuous and partial reinforcement.—The mean number of nonreinforced trials required for extinction is shown in Table 3.

An analysis of variance for the three groups, when applied to the number of trials, yields an F of 8.92 which is significant for 2 and 27 df beyond the 1% level of confidence. A comparison of the means shows that although highly significant differences are found between Groups RR and RU ($t = 3.54$, .5% level) and between Groups RR and UR ($t = 3.77$, .5% level), there exists no significant difference between Groups RU and UR ($t = .22$).

The results are, therefore, in line with the generally recognized fact that resistance to extinction is less after continuous reinforcement than after partial reinforcement (**4**).

Inequality of direction of two different paths between two regions.—Hodologically, as far as our experiment is concerned, we can speak of distinguished paths $w_{A,G}$ from a region A to a region G, in spite of the fact that the starting point A is the same as the terminal point G geographically. And at the same time, existence of two distinguished paths between A and G ($w_{A,E,G}$ and $w_{A,F,G}$) in each S's life space can be brought forth by the instructions given at the beginning of the experiment and by the tracing along each path once during the preliminary training.

TABLE 3
MEAN NUMBER OF NONREINFORCED TRIALS REQUIRED FOR EXTINCTION

Group	N	Mean	SD
RR	10	6.80	4.02
RU	10	14.60	4.63
UR	10	15.10	5.28

Direction at a region A toward a region G is different for two paths, if the two paths do not contain a common beginning step (**6**, p. 46). As for our experimental situation, then, we can make the following statement:

$$d_{A,E,G} \neq d_{A,F,G} \text{ because no path } w_{A,X} \text{ exists such that } w_{A,X} \subset w_{A,E,G} \text{ and } w_{A,X} \subset w_{A,F,G}. \quad (5)$$

Therefore, the fact that occurrence of locomotion along the path $w_{A,F,G}$ at the end of the period of extinction meant the extinction of the continuously or partially reinforced locomotion along the path $w_{A,E,G}$ shows that a psychological force $f_{A,G}$ *changed its direction* from $d_{A,E}$ to $d_{A,F}$ at the time of extinction, just as for experimental extinction of a rat in a Skinner box, extinction of a lever-pressing response was viewed as due to the changed direction of a psychological force (see formulas 3 and 4).

From the foregoing, we can conclude that *the changeability of a psychological force in direction is less after partial than after continuous reinforcement.*

So far as our experiment is concerned, however, we find no significant difference in changeability of direction of a force between the two patterns of partial reinforcement RU and UR, although Grosslight, Hall, and Murnin (**1**) found that the resistance to extinction was less after RU than after UR. Comparing their experimental conditions of partial reinforcement with ours, we can point out two main differences between them: differences in number of series in training procedure and in nature of responses partially reinforced. While their Ss had patterns RU or UR five times, our Ss had them only three times during the training period. It seems that this difference made the "patterning effect"

different in both experiments: But, the other difference, i.e., difference in nature of partially reinforced response, seems to be a more important factor. The *S*s in their study had only the very simple response of pressing a lever on each trial while our *S*s had the considerably complex serial response of moving a stylus toward a goal with as few errors as possible. Therefore, we can suppose, our *S*s were busy learning to trace correctly along the true path and were not impressed by the "patterns" of partial reinforcement.

Relative potency of a path.—It was observed not infrequently that some of the *S*s showed VTE, i.e., wavering or indecision between the green and orange paths, at the time of some trials prior to the actual extinction. Some *S*s, at such a time, talked to themselves ("Which am I to take this time, green or orange? Maybe I shall make many errors by taking the orange path."). But the fact that after the wavering many of them eventually took the original green path means that there still remained for them an "expectancy" (3) of a red light at the time of arriving at the terminal point of the path. When *S*s were wavering, however, at least sometimes it could be that

$$Po(w_{A,E,G}) < Po(w_{A,F,G}). \quad (6)$$

In such a case, therefore, we deal with a "decision" in an overlapping situation (6, pp. 34, 201–203); but such a gradual transition of the relative potency of the two paths could not be observed in all *S*s. In fact, in some *S*s, the relative potency changed suddenly and decidedly into

$$Po(w_{A,E,G}) = 0 \quad \text{and} \quad Po(w_{A,F,G}) = 1. \quad (7)$$

In any case, besides continuous or partial reinforcement, another important determinant to change direction of a

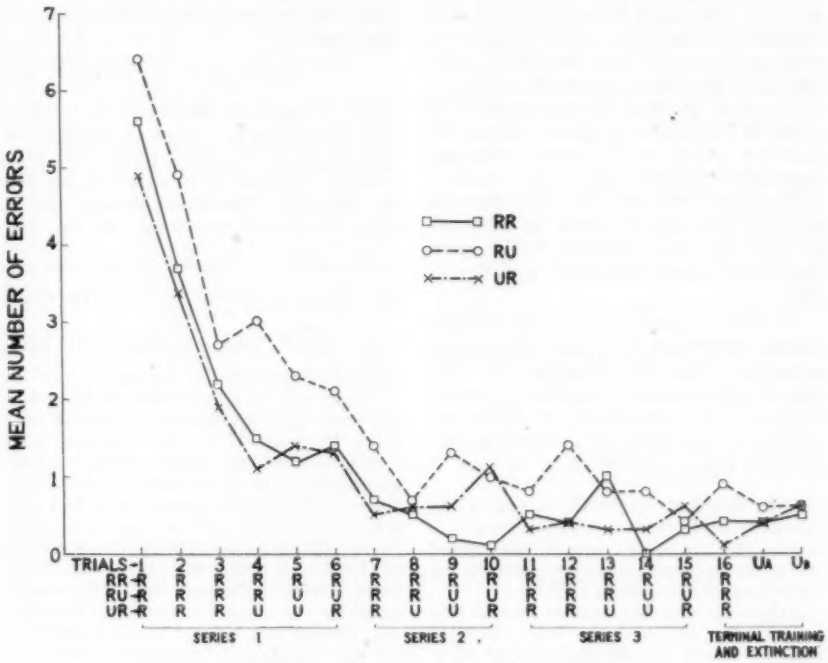

FIG. 3. Errors during the periods of training and extinction.

psychological force is relative difficulty or effortfulness of the response to be extinguished (in our case, locomotion along the path $w_{A,E,G}$) against another by which the response is to be replaced (in our case, locomotion along the path $w_{A,F,G}$).[3] As shown in Table 2, a ratio of about 1 to 3 was found in our experiment between the means of number of errors for paths $w_{A,E,G}$ and $w_{A,F,G}$.[4] We can suppose that, other factors being equal, the smaller the value of this ratio the greater the resistance to extinction, i.e., the less the changeability of a force in direction. An experiment is being planned to test this hypothesis.

"Historical causality" and "unity" of a path.—"Extinction has been a ubiquitous measure in the comparison of partial and continuous reinforcement. A large number of studies have employed this criterion of response strength in comparing the two conditions of reward and in evaluating the influence of partial reinforcement on behavior" (**4**, p. 214). By using the number of errors (buzzers in our case) as a measure of response strength, we can show error curves of the three groups as in Fig. 3. A marked decrease of the number of errors is found in each group during the training period of 16 trials, although response strength does not become asymptotic. At the same time, Groups RU and UR seem to show no influence of the six nonreinforced trials upon their error curves, compared with Group RR.

In order to know the response strength during extinction, the error curves were extended (Fig. 3). Because of individual differences in number of trials required for extinction, a procedure similar to the Vincent curve was employed: Trials for extinction in each S were divided into two parts, i.e., the former and latter halves (U_A and U_B in Fig. 3)[5] and means of number of errors per trial were calculated for both halves of each group. Since the mean number of errors for all Ss on the last trial of green-path response, i.e., the trial just prior to extinction, was .7, this value and values of U_B in the three groups indicate a remarkably great response strength. Therefore, the response strength just prior to extinction as measured by number of errors was at the highest level.

The fact that number of errors gradually decreased means a step-by-step change in cognitive structure of the psychological field (**6**, pp. 57–61) and an increased degree of "unity" of the path $w_{A,E,G}$ crossing several regions. As Lewin stated, "The development of automatism during learning decreases the relative potency of the single steps. A routine path shows, therefore, a relatively high degree of unity" (**6**, p. 35). This contention, when applied to our case, means

$$uni(w_{A,E,G}) = F(N) \qquad (8)$$

where N indicates the number of trials. This is nothing less than an example of "historical statements"[6] (**6**, p. 165) concerning a "historical process" (**6**, pp. 81, 82, 130, 133), "historical change" (**6**, p. 132), "historical problem" (**6**, pp. 82,

[3] The fact that "bar load" is one of the determinants of resistance to extinction in case of rats (**8**) is a special case of this general statement, i.e., case of infinite difficulty or effortfulness of the replacing response.

[4] There was a marked tendency for each S to underestimate the number of errors (buzzers) made by himself, but this fact does not affect present interpretations of our results.

[5] In performing this division, we made the the former half contain $\frac{n+1}{2}$, and the latter $\frac{n-1}{2}$, when S's number of nonreinforced trials required for extinction (n) was odd.

[6] It seems to be correct that field theory has put stress upon "contemporary causation" (**3**, p. 15) or "principle of contemporaneity" (**7**, p. 45), because of its taking a serious view of statements concerning the distribution of forces at a given time to explain any types of behavior. But, if field theory is to deal with the relation between the situations at two different moments, i.e., learning, it must make place for the "historical process." The term "historical process" should not only mean the two types of learning, i.e., "problem solving" and "habit formation," but also "extinction," "spontaneous recovery," and other phenomena related to learning.

134), or a "historical causality" (5, p. 30). The fact that a considerably small number of errors can be found in the U_B in each group and on the last trial for extinction in all *S*s shows that the degree of unity $uni(w_{A,E,G})$ was relatively high even just before complete extinction.

Further, the fact that the green-path response underwent extinction in spite of its considerably high level of unity seems to mean that extinction has no bearing upon the unity of response: In terms of reinforcement theory, there seems to be no relation between habit strength and extinction.[7] The distinction between *learning* and *performance* which was made originally by Tolman (10, p. 364) may be applied to the present case of extinction: Extinction seems to be due to some change in drive, that is, *D* in an equation ${}_sE_R = f(D) \times f({}_sH_R)$. This change, however, seems to be qualitative in nature, rather than quantitative. We view this qualitative change as *change in direction of a psychological force.* Only by using such a *cognitive* concept as direction can we deal adequately with the extinction phenomena of our experiment.

Summary

In a situation where extinction of continuously or partially reinforced serial response was indicated by the change to an alternative response for the same goal, resistance to extinction for thirty university students was decidedly smaller after continuous reinforcement than after partial reinforcement. No significant difference was found between RU and UR as different patterns of partial reinforcement. As measured by the number of errors, "unity" (i.e., "habit strength" in terms of reinforcement theory) of a response to be extinguished was nearly at the highest level right up to time of complete extinction. These results seem to indicate that extinction must be due to the change in direction of a force rather than to the decrease in habit strength. Therefore, resistance to extinction may be defined as a function of changeability of a force in direction.

[7] Latency has been used as a measure of resistance to extinction. According to the author's viewpoint, however, latency in acquisition has not the same psychological meaning as it has in extinction, even though they are quite similar in their *operational* or *empirical* characteristics. Generally speaking, *phenotypical or directly observable similarities do not always guarantee genotypical or dynamic ones.* Latency in extinction, then, cannot be viewed as an adequate measure for habit strength in extinction. The increase of latency in extinction could be viewed as due to something like a "wavering" or "indecision" stated above, i.e., *undecided direction* of a psychological force, rather than due to the decrease of habit strength. Latency in acquisition could be viewed as due to the *cognitively unstructured* psychological field.

References

1. Grosslight, J. H., Hall, J. F., & Murnin, J. Patterning effect in partial reinforcement. *J. exp. Psychol.*, 1953, **46**, 103–106.
2. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
3. Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
4. Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
5. Lewin, K. *Principles of topological psychology.* New York: McGraw-Hill, 1936.
6. Lewin, K. The conceptual representation and the measurement of psychological forces. *Contr. psychol. Theor.*, 1938, **1**, No. 4.
7. Lewin, K. *Field theory in social science.* New York: Harper, 1951.
8. Mowrer, O. H., & Johns, H. M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33**, 369–386.
9. Sartain, A. Q. The conception of the true path and efficiency in maze learning. *J. exp. Psychol.*, 1940, **26**, 74–93.
10. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

(Received May 25, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

FLICKER FUSION MEASUREMENTS AND ANXIETY LEVEL[1]

SANFORD GOLDSTONE

Duke University[2]

Krugman (3) reported the results of an exploratory study which suggested that sensitivity to a flickering light (flicker fusion threshold) was inversely related to anxiety level. The present study represents an empirical investigation of the relationship between flicker fusion measurements and anxiety level in order to provide further information on anxiety as it relates to performance on a sensory discrimination task. The flicker fusion method has a long history with a current emphasis upon the relationship between flicker acuity and nonretinal conditions. Recent investigations summarized by Landis (4) and Simonson and Brožek (5) demonstrated a relationship between flicker sensitivity and fatigue, impaired oxygen circulation, ingestion of central and autonomic nervous system stimulants and depressants, ophthalmologic pathology in any part of the visual pathway, central nervous system pathology, cardiovascular function, and possibly anxiety.

The flicker fusion measures included in this study, and related to anxiety level were the following: (*a*) flicker fusion threshold (FFT), the measure of sensitivity to flicker, and (*b*) slope, the measure of intrasubject variability or precision of judgment. Anxiety is here defined in terms of those complaints or symptoms usually associated with the psychiatric diagnosis *anxiety reaction*.

Method

Apparatus.—The present study employed an electronically activated "glow modulator" tube which illuminated a 5-mm. circular test patch mounted in a black box at 1° visual angle. The patch appeared as white light at a brightness of .35 mL and the light-dark ratio used was 60:40 over the entire frequency range. Calibration was constantly checked by a neon zero set. The examination room was kept dimly illuminated by a shielded 60-w. lamp mounted in a far corner of the room.

Subjects.—Seventy-four *S*s meeting specific criteria (2) and selected serially from the patients and their relatives in the Duke out-patient psychiatric clinic were separated into the following groups according to the anxiety criteria to be described below: 33 high-anxiety patients, 18 low-anxiety patients, 2 high-anxiety normals, 21 low-anxiety normals, 35 total high anxiety, 39 total low anxiety.

Procedure.—All *S*s received a 3-min. adaptation period followed by a method of limits flicker fusion test. Alternating runs of ascending and descending flicker rates were presented. The duration of stimulus presentation was 1.5 sec. and the step interval was .5 cycle per second. Three blocks of five ascending runs alternating with five descending runs were used, separated by a 30-sec. rest. The percentages of reports of *flicker* for each frequency in the last 20 trials were plotted on normal probability paper (1) on the assumption that these values were normally distributed, and a straight line was visually fitted. The FFT or frequency at which *S* reported flicker 50% of the time, and slope or *SD* of *S*'s judgments were derived from these plots. The following flicker fusion measures were studied for each group: (*a*) FFT, or sensitivity to flicker; (*b*) slope or precision of judgment (intra-*S* variability); (*c*) decline in FFT, which was defined as the difference in flicker fusion threshold between the first ten and

[1] Based on a dissertation submitted to the Department of Psychology of Duke University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Appreciation is due Dr. L. D. Cohen for supervising this project and Drs. J. Goldsmith and H. Lowenbach for facilitating the implementation of this investigation.

[2] Now at Department of Psychiatry, Baylor University College of Medicine, Houston, Texas.

TABLE 1

THE MEAN FFT, SLOPE, FFT DECLINE, AND SLOPE DECLINE IN CYCLES PER SECOND

Groups	*N*	FFT		Slope		FFT Decline		Slope Decline	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Patients									
High anxiety	33	31.7	3.62	1.2	0.39	2.4	1.33	0.1	0.55
Low anxiety	18	36.9	1.86	0.8	0.32	0.3	0.78	0.0	0.33
Normals									
High anxiety*	2	32.0		0.8		1.6		0.4	
Low anxiety	21	37.1	1.55	0.7	0.22	0.2	0.87	0.2	0.41
Total									
High anxiety	35	31.7	3.17	1.1	0.39	2.3	1.31	0.1	0.50
Low anxiety	39	37.0	1.68	0.7	0.26	0.3	0.26	0.1	0.37

* This group was not used for group comparisons due to the limited *N*.

last ten runs of the flicker test (a decline in FFT would reflect a reduced sensitivity to flicker associated with continued exposure to the flicker test); (*d*) decline in slope, which was defined as the difference in slope between the first ten and last ten runs of the flicker test (a decline in slope would indicate an increased precision of judgment associated with continued exposure to the flicker test).

Anxiety criteria.—In order to separate our *Ss* into high- and low-anxiety groups, two criteria were employed. The first was Lorr's anxiety-tension rating scales[3] on which *E* rated *S* for irritability, tension, anxiety, and frustration tolerance on six-point scales, and sleep complaints on a four-point scale. The second was a check list developed at Duke Hospital (**2**) on which *S* designated those symptoms applying to him. Cutoff points were established in preliminary investigation, and when the two measures agreed upon the placement of an *S* in either a high- or low-anxiety group, that *S* was included in this study. These measures were checked against psychiatric diagnoses and were found to successfully select those patients diagnosed anxiety reaction for the high-anxiety group, thus providing positive evidence relevant to their validity.

RESULTS

Table 1 shows the means and *SD*'s of the FFT, slope, FFT decline, and slope decline in cycles per second for the six groups. No significance tests were performed involving the high-anxiety normal group because of the limited number of *Ss*. The differences between groups were examined by the *t* test. In each case the assumption of homogeneity of variance was tested and when necessary the *t* test was modified accordingly.

With regard to the mean FFT one observes a lower flicker threshold for those groups designated high anxiety than for the groups designated low anxiety. These group differences in all cases appear reliable at better than the .005 level of confidence and indicate a reduced sensitivity to flicker in the high-anxiety groups.

With regard to the mean slope or *SD* of *Ss*' judgments, one notes a higher index of scatter for those groups designated high anxiety than for those designated low anxiety. The group differences in all cases appear reliable at better than the .005 level of confidence and indicate more variable judgments on the part of *Ss* in the high-anxiety groups.

For the mean FFT decline it may be observed that the groups designated high anxiety have a greater drop in sensitivity to flicker than groups designated low anxiety. These group

[3] Lorr, M. Personal communication, 1952.

differences in all cases appear reliable at better than the .005 level of confidence and reflect a greater reduction in flicker acuity associated with continued exposure to the flicker test in the high-anxiety groups.

No significant difference between groups on mean slope decline is observed although a slight trend in the direction of decreased variability of judgment with practice is noted for most groups.

Discussion

These data suggest a reduced sensitivity to flicker, decreased precision of judgment, and a decline in sensitivity associated with continued exposure to the flicker test in groups designated high anxiety as compared with groups designated low anxiety. It would appear then that this study verifies Krugman's earlier suggestion that a lower flicker threshold might be expected to accompany anxiety.

These findings raise the question of whether we are dealing with a general impairment of sensory discrimination associated with anxiety or whether this reduced flicker acuity is perhaps more locally and physiologically determined. The literature (**4, 5**) describes many physiological conditions which are, like anxiety, associated with reduced flicker sensitivity.

While the present study demonstrated certain relationships using the extremes of anxiety level, it would appear to be important to study the extent to which these relationships hold between high- and low-anxiety *Ss* within a normal population in order to explore the limits of sensitivity of this technique.

Summary

This study involved 33 high-anxiety patients, 18 low-anxiety patients, 2 high-anxiety normals, and 21 low-anxiety normals. A flicker fusion test was administered to all *Ss* and the following reliable differences were found:

1. Those groups designated high anxiety had a lower FFT than those groups designated low anxiety, suggesting reduced sensitivity to flicker associated with high anxiety.
2. Those groups designated high anxiety had a greater intraindividual variability of judgments than those groups designated low anxiety.
3. Those groups designated high anxiety had a greater decline in sensitivity to flicker (FFT) associated with continued exposure to the flicker test than those groups designated low anxiety.

References

1. Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
2. Goldstone, S. Flicker fusion measurements and anxiety level. Unpublished Ph.D. thesis, Duke Univ., 1953.
3. Krugman, M. Flicker fusion frequency as a function of anxiety reaction; an exploratory study. *Psychosom. Med.*, 1947, **9**, 269–272.
4. Landis, C. Something about flicker fusion. *Sci. Mon.*, 1951, **73**, 308–314.
5. Simonson, E., & Brožek, J. Flicker fusion frequency: background and applications. *Physiol. Rev.*, 1952, **32**, 349–378.

(Received May 27, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

CONSTANT VELOCITY TRACKING AS A FUNCTION OF S'S HANDEDNESS AND THE RATE AND DIRECTION OF THE TARGET COURSE[1]

DAVID A. GRANT AND NOEL F. KAESTNER

University of Wisconsin

The present investigation is concerned with the relations between hand dominance of *S*, target direction, and target speed in a unidimensional tracking task. Since the target motion was confined to one dimension, horizontally presented, a comparison of right-to-left tracking with left-to-right tracking is possible. Chapanis, Garner, and Morgan (**2**) have discussed the influence of directional factors in positioning responses. Here, the directional factor is examined in a tracking situation. Handedness has been investigated by Simon, DeCrow, Lincoln, and Smith (**6**) using a slow but complicated target movement. They found no differences between right- and left-handed *S*s, tracking with either preferred or nonpreferred hand.

In an earlier study in the series (**4**), *S*s were asked to compare the two directions of target motion with regard to difficulty. Their responses indicated that they thought that there was a difference in difficulty. But there was little accord on which direction was more difficult. The majority of *S*s offered variations of two main hypotheses. The more frequent explanation postulated that movements of the hand control are steadier in one direction than the other. Among those who offered this motor interpretation, there was disagreement between left- and right-handed *S*s as to which direction of target motion was easier to track; i.e., left-handed *S*s reported left-to-right tracking was easier while right-handed *S*s favored the opposite direction. When we incorporate this handedness difference into a general motor hypothesis, the fact emerges that *S*s seem to prefer to track targets which require movements directed medially from their dominant side.

A second interpretation of difficulty differences due to direction assumed that perception was more accurate where the target moved from left to right. Some extended this hypothesis by attributing this perceptual difference to the effect of reading habits. It will be noted, however, that saccadic movements are involved in reading, whereas tracking involves pursuit movements of the eye (moving fixation). According to Woodworth (**7**), saccadic movements are associated with pursuit movements only in: (*a*) correcting for lag during the pursuit phase of movements maintained for a considerable duration; and (*b*) pursuit movements which attempt to follow especially complex external motion that may cause saccadic movements to appear. Since neither condition *a* nor *b* prevailed in the study to be described, it appears reasonable to discount a "reading habit" interpretation of directional proficiency if it should emerge in the

[1] This research was supported in part by the United States Air Force under Contract Number AF 33(038)-23294 monitored by the Skill Components Research Laboratory, AF Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.

data The decision does not, however, rule out all perceptual hypotheses for directional differences.

In addition to investigating the directional variable, we examined the relation between target speed and tracking proficiency. Besides defining the speed-accuracy relationship, the design lends generality to the conclusions reached regarding the directional factor. The directional factor was investigated at four levels of this second parameter, speed of target motion. In addition to the above three variables (target direction, handedness, and rate of target movement), *Ss*' past laboratory experience with tracking was maintained orthogonal to the above variables. Half of the *Ss* employed here had participated in an earlier experiment. A period of three months intervened between the two studies.

Method

Apparatus.—A unidimensional tracking device described in the earlier study (**4**) was employed. The only physical alteration was in the hand control. The vertical joy-stick control was replaced by a control of greater mass. It consisted essentially of a 10-in. wheel to which a cylindrical spinner-type handle, $\frac{1}{16}$ in. in diameter and $2\frac{1}{2}$ in. long, was affixed. Both wheel and spinner were nearly friction free. The total angular movement of the control wheel was 80° which effected a $4\frac{1}{2}$-in. excursion of the cursor, giving a constant control-display sensitivity of 18–22° control movement for each inch of display movement. The total hand movement amounted to a little less than 8 in., and, since the spinner handle was at the top of the wheel and projected toward *S*, movements of handwheel and cursor from right to left involved extension of forearm and upper arm in right-handed *Ss* and flexion of these members in left-handed *Ss*. In addition, the scoring system was altered. Instead of recording time on target with Standard Electric timers, counters were installed to record the number of tenths of a second *S* was continuously on target. This scoring device was noncumulative in that two target contacts, each of less than .1 sec., did not cumulate to give one digital advance of the counter.

The stimulus display appeared through a horizontal slit, $4\frac{1}{2}$ in. $\times \frac{1}{8}$ in., cut in the front of the gray steel cabinet before which *S* sat during tracking. Behind the slit the target track was propelled upward from the bottom by an Esterline-Angus chart drive. The target was about $\frac{1}{32}$ in. wide, but the scoring area was about $\frac{1}{16}$ in. wide. Variations in target speed resulted from differences in the slopes of the target lines since the paper itself moved at a constant speed. The *S*'s task was to place the cursor on the target and keep it there as much as possible. When *S* was on target, a thyratron circuit picked up the contact through the cursor, the electrograph target line, and the stable contact below the slit. It was this circuit that supplied the input to the counter system described above.

Procedure.—Eight experimental conditions evolved from the orthogonality of directions and the four target rates. Every *S* appeared once in each condition; only the order of the conditions varied between *Ss*. Each condition or series consisted of four trials, each 75 sec. in duration. The targets within a series of trials had uniform direction and velocity; it was the particular values of these qualities which distinguished a given condition from the others. On Day 1, after being instructed as to what he was to do, *S* received his first four series of trials. Rests between series and trials were 150 and 45 sec., respectively. The *Ss* returned to complete their training on Day 2 and received the last four series at that time. The sequence of the eight conditions over the two days for a given *S* was randomly determined.

Targets.—In the present study the targets on a single series of trials moved with constant velocity from left to right or right to left. The horizontal extent of target motion was $4\frac{1}{2}$ in. under all conditions. The targets were obtained by drawing a series of parallel, evenly spaced lines, extending diagonally across the Esterline-Angus chart. The rates of target movement depended upon the slope of the diagonal lines, because the paper passed the target slit at a constant rate. The four target rates were 5.33 in./sec., 2.67 in./sec., 1.78 in./sec., and 1.33 in./sec. Owing to the constant trial length, more fast target lines than slow target lines appeared on a single trial.

Subjects and design.—Twenty-four male *Ss* were paid to participate in the experiment. Sixteen were right-handed; eight of these had had prior practice in an earlier experiment, while the others were naïve with regard to the apparatus. Among the eight left-handers, half were naïve. The eight *Ss* of each group (eight experienced right dominant, eight naïve right dominant, and eight mixed experience, left dominant) formed the rows of an 8×8 latin

square. There was one replication of this square for each group of eight *Ss*. The other side of the square (columns) represented successive practice series. The latin letters corresponded to the eight experimental conditions. Replication of the square permitted evaluation of the sequence factor, and served to alert *E* to any possible unique patterns of interactions in the square employed.

Results

Analyses of individual groups.—Figure 1 shows the course of improvement for each of the three handedness groups. The number of target hits per *S* during each 75-sec. trial is plotted against successive series of trials (each of 75-sec. duration). The three curves indicate that the learning trends are quite similar for all groups, showing a regular learning process with few inversions. The inexperienced right-handed group showed the greatest improvement over their initial series performance (44%) and the left-handed group improved the least (20%). Tests of significance on these series to series changes will be discussed separately for each handedness group below.

Results of experienced right-handed Ss.—Figure 2 presents the performance scores as a function of target speed separately for the three handedness groups with the target direction as the parameter. The data of the experienced right-handed *Ss* are given in Fig. 2A. Tracking performance (in terms of hits per *S* per 75-sec. trial) falls off linearly as a function of target speed. With high target speed, the right-handed *Ss* are superior in tracking targets which move from right to left.

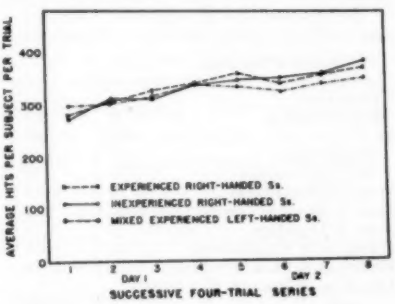

FIG. 1. Practice effects on successive four-trial target series. Each point is based on 8 *Ss*, no two of which tracked the same target in the given series. The *S*'s handedness and tracking experience are the parameters.

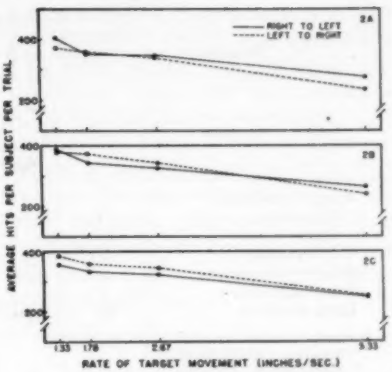

FIG. 2. Tracking accuracy as a function of target speed. Data are plotted for experienced right-handed *Ss* (2A), inexperienced right-handed *Ss* (2B), and left-handed *Ss* of mixed tracking experience (2C). Direction of target motion is represented as the parameter in each subfigure.

The data of Fig. 1 and 2A were analyzed by the usual latin-square procedures. Table 1 presents the summary of the analysis for experienced right-handed *Ss*, where it is found that Target Rates, *Ss*, and Practice Effect all constitute significant sources of variation. The dominant trends of Fig. 1 and 2A are therefore found to be significant.

Results of naïve right-handed Ss.—The effect of target rate on tracking performance for the naïve right-handed *Ss* is presented in Fig. 2B. As with the experienced *Ss*, there is a linear reduction in tracking score with increased target speeds. Neither tar-

TABLE 1

SUMMARIES OF ANALYSES OF VARIANCE OF PERFORMANCE SCORES

Source of Variation	*df*	*F*		
		Experienced Right-Handed *Ss*	Inexperienced Right-Handed *Ss*	Left-Handed *Ss*
Target conditions	7	13.53**	12.77**	20.66**
Direction	(1)	3.52	0.01	1.95
Rate	(3)	29.37**	28.43**	45.39**
Interaction	(3)	1.03	1.31	2.16
Subjects (sequences)	7	9.90*	14.93**	9.56*
Practice effects	7	4.45*	5.43*	3.35*
Error variance	42	1,776†	1,967†	1,047†

* Significant at 1% level.
** Significant at .1% level.
† Mean squares.

get direction was uniformly superior for the inexperienced right-handed *Ss*.

The pertinent significance tests for inexperienced right-handed *Ss* are given in Table 1, where it is shown that Target Rates, Practice, and *Ss* produce significant effects. In general, the significance tests closely parallel those for the experienced right-handed *Ss*.

Results of left-handed Ss.—The effects of target speed and direction on tracking scores of the combined groups of naïve and experienced left-handed *Ss* are shown in Fig. 2C. The left-handed *Ss* exhibited about as much preference for the left-to-right target direction as did the experienced right-handed *Ss* for the right-to-left target direction. The main characteristic of Fig. 2C, however, is the effect of target speed on tracking scores. Again, the reduction in tracking scores was approximately linearly affected by increased target speed. The significance tests for the left-handed *Ss* are given in Table 1, where it is again shown that the three main significant sources of variation were Practice Effect, Target Rate, and individual differences between *Ss*. The improvement sequence (Practice Effect) was very irregular as shown in Fig. 1, but the *F* of 3.35 was significant. Since half of the left-handed *Ss* were inexperienced and half were experienced, it was possible to test the factor of previous experience on overall tracking score. The difference between the experienced and inexperienced groups was found to be insignificant as was also the case for the difference between the experienced and inexperienced right-handed *Ss*.

Analysis of the pooled groups.—The data of all three handedness groups

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF POOLED DATA FROM ALL HANDEDNESS GROUPS

Source of Variation	*df*	*F*
Target rate	3	62.41**
Target direction	1	0.05
Ss	23	7.05**
Handedness	1	0.20
Ss within groups	22	7.30**
Rate × Handedness	3	0.13
Rate × *Ss* within groups	66	1.03
Rate × Direction	3	1.58
Direction × Handedness	1	6.71*
Direction × *Ss* within groups	22	0.51
Direction × Rate × Handedness	3	0.45
Direction × Rate × *Ss* within groups	66	2,483†

* Significant at 5% level.
** Significant at .1% level.
† Mean square.

were combined in order to permit evaluation of certain sources of variation not previously examined. When the three groups were compared on over-all tracking ability, it appeared that most right-handed *Ss* were superior to the left-handed *Ss*. Even the inexperienced right-handed *Ss* seemed to track better than the left-handed *Ss*, half of whom had had prior training. The statistical analysis of the pooled data is summarized in Table 2, and when the Handedness mean square was tested with the *Ss* Within Groups mean square, the Handedness *F* was insignificant, being less than one.

When the handedness groups were compared with respect to the directional factor, two interesting relations emerged. First of all, the directional component of target differences tended to become more important as *Ss*' tracking experience increased; i.e., the inexperienced right-handed *Ss* showed no preference for either direction, the left-handed group, intermediate in previous tracking experience, showed a tendency to favor left-to-right target motion, and the experienced right-handed *Ss* showed a distinct preference for the right-to-left target direction. Individually these differences were not statistically significant. When the right-to-left scores of the two right-handed groups were compared with those of the left-handed group, the right-handed groups were superior ($F = 7.8$, $p < .01$). This difference combines the preference factor of right-handed *Ss* for right-to-left tracking and the general superiority of right-handed *Ss* over left-handed *Ss*. There was no significant difference between handedness groups in left-to-right tracking. When these over-all observations were put to a single statistical test, consisting of an evaluation of the interaction between Handedness and Target Direction, the *F* was significant. This mean square was tested against the mean square for the Direction × *Ss* Within Groups interaction. All other *F*'s in Table 2 involve the mean square of the Direction × Rate × *Ss* Within Groups interaction as the error estimate. The only significant difference between handedness groups was in the preference for the different target directions.

Discussion

The effects of single variables in the present experiment were similar to those reported in previous studies. Target rate (**3, 5**) and practice effects introduced significant variation in the performance measure. On the other hand, neither handedness (**6**) nor the direction of target motion had an effect individually upon the tracking performance measures, but there was a combined effect or interaction of handedness and target direction that was statistically significant.

The Handedness × Target Direction interaction consisted of a superiority of the right-handed *Ss* on right-to-left target motions and a superiority of the left-handed *Ss* on left-to-right target motions. Such a result has not been reported in the previous tracking literature, and the present experiment was not designed to produce analytical interpretations of this interaction. It seems unlikely that perceptual factors could account for the difference in difficulty of the different target directions, because both right- and left-handed *Ss* were used to reading from left to right. Furthermore, neither handedness group was aided or handicapped in viewing the target, because the handwheel was well below the target display. It would appear that the difference is due largely to a motor factor related to the observation of Brown, Knauft, and Rosenbaum (**1**) that positioning movements are more accurately made with a thrusting or extending motion of the hand and arm than by pulling or flexing motions. In

tracking a right-to-left target, the right-handed *Ss* were pushing the handwheel handle and when they tracked a left-to-right target they had to pull the handle. The left-handed *Ss*, in tracking the left-to-right target motions, pushed the control stick, whereas in tracking right-to-left target motions they had to pull. The above interpretation is ex post facto and requires further investigation.

Summary

The effect of handedness, target direction, and target speed in a unidimensional tracking task was investigated in a latin-square experiment involving 24 *Ss*. Each *S* tracked for 5 min. under each of eight experimental conditions. There were appropriate rest periods after each 75 sec. of tracking. The eight experimental conditions consisted of the factorial combinations of (*a*) right-to-left target motions vs. left-to-right target motions, and (*b*) four target rates, 1.33 in./sec., 1.78 in./sec., 2.67 in./sec., and 5.33 in./sec. Eight of the *Ss* were right-handed and had had experience on the unidimensional tracking task. Eight other right-handed *Ss* were inexperienced. Of the eight left-handed *Ss* used, half were experienced and half were inexperienced. The principal findings of the experiment were as follows:

1. There was a clear cut and significant practice effect for all *Ss*.

2. Tracking performance scores fell off linearly as a function of the target speed for all groups of *Ss*.

3. Neither handedness nor target direction, singly, contributed to the significant variation in the experiment, but they had a combined effect or interaction that was statistically significant. The right-handed *Ss* tracked the right-to-left targets better, and the left-handed *Ss* tracked the left-to-right targets better. This finding was interpreted in terms of a motor factor dependent upon the superiority of positioning movements involving a pushing or thrusting as opposed to a pulling motion.

References

1. Brown, J. S., Knauft, E. B., & Rosenbaum, G. The accuracy of positioning movements as a function of their direction and extent. *Amer. J. Psychol.*, 1948, **61**, 167–182.
2. Chapanis, A., Garner, W. R., & Morgan, C. T. *Applied experimental psychology*. New York: Wiley, 1949.
3. Hick, W. E., & Bates, J. A. V. *The human operator of control mechanisms*. London: Ministry of Supply, 1950.
4. Kaestner, N. F., & Grant, D. A. Transfer of training in tracking as a function of the predictability of unidimensional target courses. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1954.
5. Searle, L. V., & Taylor, F. V. Studies of tracking behavior. I. Rate and time characteristics of simple corrective movements. *J. exp. Psychol.*, 1948, **38**, 615–631.
6. Simon, J. R., DeCrow, T. W., Lincoln, R. S., & Smith, K. U. Effects of handedness on tracking accuracy. *Percept. mot. Skills Res. Exch.*, 1952, **4**, 53–57.
7. Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received May 24, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

THE RELATION BETWEEN THE RATE THRESHOLD FOR THE PERCEPTION OF MOVEMENT AND LUMINANCE FOR VARIOUS DURATIONS OF EXPOSURE [1]

H. W. LEIBOWITZ [2]

University of Wisconsin

By experimental variation of luminance and duration of exposure, it has been possible to systematize a large body of visual discrimination data (**7, 10**). The purpose of the present experiment is to determine velocity thresholds for the perception of movement, with exposure time as a parameter, over a wide range of adapting luminances. Such data should provide functional descriptions of movement discrimination relationships and permit analysis of the mechanisms subserving the perception of motion.

Experimental studies of real movement have utilized a number of procedures and criteria for the detection of motion. The *rate threshold* is the minimum discriminable target velocity (ds/dt) (**5**). In the present study, the threshold velocity for a constant duration of exposure has been investigated and will be referred to as the *isochronal threshold velocity*. Velocity thresholds obtained when displacement of the test object is constant may be designated as the *isometric threshold velocity*, a necessary distinction since rate thresholds may be investigated with either time or displacement as parameters. In all cases, the targets will be presented at constant velocities ($d^2s/dt^2 = 0$).

[1] This research was supported in part under Contract AF 18(600)-54 between the University of Wisconsin and the Psychology Branch, Aero Medical Laboratory, Directorate of Research, Wright Air Development Center, Air Research and Development Command, Wright-Patterson Air Force Base, Ohio.

[2] The author wishes to express his appreciation to Dr. D. A. Grant for his help in the preparation of the manuscript and to acknowledge the assistance of Mr. J. F. Lomont in obtaining the data.

METHOD

Apparatus.—The targets used in this study are white rectangular squares each subtending a visual angle of 15 min. of arc at a viewing distance of 90.6 in. (.043 diopter accommodation). These squares are pasted to a black belt $\frac{1}{8}$ in. in width and are spaced so that the angle between the centers of adjacent targets is 45 min. of arc. The belt carrying the targets appears in a rectangular opening 5 in. by $\frac{1}{8}$ in. (see Fig. 1B).

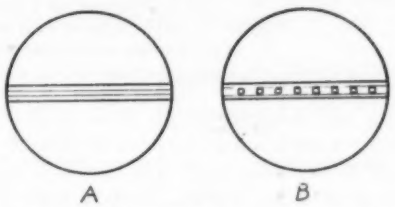

FIG. 1. Schematic diagram of the *S*'s view of the apparatus. A, between trials; B, during exposure of test object.

The area in back of the belt and the frame from which the rectangular opening is cut are painted "dead" black. The *S* is seated in a dark room and views the targets monocularly with the natural pupil through a black tube, the end of which provides a circular field stop limiting the view of the frame to a circle subtending a visual angle of 3.2 deg. of arc. The *S*'s view of the belt is controlled by means of a shutter located just in front of the belt and activated by means of a solenoid and a Microflex mechanical interval timer. With the shutter raised *S* sees the belt as described; with the shutter lowered, he sees a white rectangular strip $\frac{1}{16}$ in. by 5 in. which aids in maintaining orientation of the eye in the direction of the targets (Fig. 1A).

Movement of the belt is provided by means of power transmitted from electric motors through

a Metron variable speed reducer. The Metron provides a speed variation of 36 to 1. By use of several motors and a fixed speed reducer, velocities from .1 min. of arc per sec. to 76 min. of arc per sec. can be obtained.

Illumination is provided by two 300-w. reflector flood lamps operated on 110 v. AC, directed toward the target areas. Luminance ripple resulting from the alternating current, determined with the aid of a photocell and oscilloscope, is less than 4%. The luminance of the targets as seen by the *S* is controlled by Wratten neutral tint filters placed in the viewing tube.

Procedure.—Thresholds for the detection of motion were determined by increasing the velocity of the belt carrying the targets in steps bearing the ratio of 1.35 to 1.00 from subliminal values until a value at which motion was perceived twice in succession. Since in the present experiment the number of reference stimuli in the visual field was purposely reduced to a minimum, the experimental arrangement was favorable for the appearance of illusions of motion. For this reason, only ascending series of velocities were used, it having been discovered in preliminary experimentation that descending series would induce illusory motion which would persist even when the targets were stationary. In addition, direction of target motion was varied randomly to the right and to the left, and *S* was required to indicate the direction of movement. A third procedure, designed to eliminate illusory motion, was to show *S* the stationary targets at the beginning of each series of threshold determinations.

The procedure for a given experimental session, lasting from 1 to 1.5 hr., was as follows: After dark adapting for 10 min., *S* was shown the stationary targets at the lowest luminance level (.008 mL) for the duration of exposure which was to be used for that day. Threshold velocities were then investigated at that luminance by determining from two to four pairs of thresholds, each pair consisting of the means of movement thresholds for motion to the right

TABLE 1

ISOCHRONAL THRESHOLD VELOCITIES IN MINUTES OF ARC PER SECOND AS A FUNCTION OF LUMINANCE AND EXPOSURE TIME

Targets were white squares, each subtending 15 min. of arc spaced 45 min. of arc between centers and viewed against a black background.

S	Luminance (Millilamberts)	Exposure Time (Seconds)				
		1/8	1/4	1	2	16*
B	.008	—	55.70	8.00	7.20	0.66
	.016	36.00	15.20	5.10	2.31	0.39
	.05	14.00	7.00	2.73	1.30	0.26
	.16	8.30	6.80	1.75	1.00	0.50
	.5	6.10	3.90	1.63	0.80	0.26
	5	3.80	2.80	1.74	0.80	0.48
	50	3.80	2.90	1.50	0.59	0.44
	500	2.90	2.40	1.74	0.63	0.30
P	.008	—	21.90	2.74	1.76	0.51
	.016	50.90	8.60	1.46	1.11	0.35
	.05	23.60	6.80	1.24	0.74	0.36
	.16	27.00	3.50	0.83	0.54	0.31
	.5	16.10	4.10	0.57	0.34	0.28
	5	7.70	3.00	0.38	0.32	0.20
	50	5.00	3.00	0.31	0.32	0.25
	500	4.30	3.00	0.37	0.32	0.22
W	.008	27.00	25.10	6.80	4.20	0.33
	.016	20.40	15.90	3.48	3.00	0.36
	.05	18.90	11.80	2.73	1.73	0.32
	.16	15.20	6.20	2.33	1.11	0.31
	.5	10.00	4.60	2.72	0.63	0.31
	5	5.00	3.40	2.03	0.59	0.28
	50	3.30	3.00	1.35	0.52	0.32
	500	3.20	2.80	0.79	0.50	0.23

* The full 16-sec. interval had not always expired when *S* reported movement. See text for details.

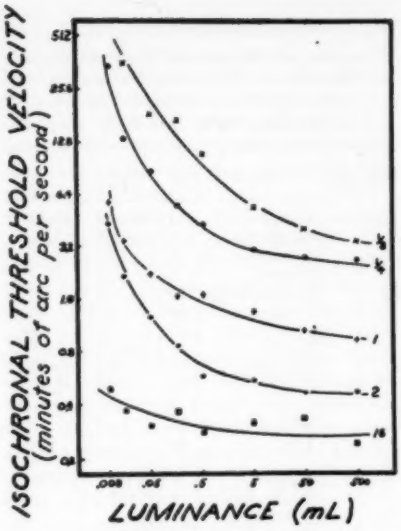

FIG. 2. Isochronal threshold velocity as a function of luminance with duration of exposure as parameter.

and to the left. This procedure was repeated for seven successively higher luminance levels of .016, .05, .16, .5, 5, 50, and 500 mL. Five exposure times of ¼, ½, 1, 2, and 16 sec. were employed as parameters and three experimental sessions were devoted to each exposure time. The number of thresholds which could be determined in a given experimental session depended upon the exposure time. Because of the length of time required for a single judgment at the longest exposure, four thresholds were obtained at each luminance level while eight thresholds could be determined at the shortest duration of exposure. Either six or eight thresholds were obtained at the intermediate durations depending upon the rapidity with which the experiment could be performed.

At the 16-sec. exposure, a special problem arose. The *S*s would often report motion before the termination of the 16-sec. interval. At times, motion was perceived within 4 sec. while in other cases the full 16 sec. were required. The average time was about 13 sec. Presumably, the threshold velocity might always be determined for the full 16 sec. by increasing the velocity in smaller steps than were used here. Since this was not feasible in the present study, *S*s were instructed to report immediately when motion was perceived. The shutter was then closed and the time utilized by *S* before signalling motion was recorded. In view of this, the 16-sec. duration should be more precisely specified as a *maximum* of 16 sec. No relation was observed between the time required to see motion and the velocity threshold obtained at the 16-sec. exposure.

Another problem which arose concerns the variability of *S*s' performance and the reduction thereof as they became more experienced. At first, *S*s' responses were variable. It was for this reason that the criterion of two correct responses at a given velocity was adopted. With experience, this variability was reduced and the need for the two-response criterion eliminated. It was retained, however, for the sake of uniformity of procedure. Before the data of the present paper were obtained, *S*s were given from five to ten practice sessions on the apparatus (these sessions served to provide information to be used in planning the study and modifying the apparatus). At the beginning of the main experiment, the three *S*s whose data are reported here had presumably been stabilized. This inference is based on the similarity of data obtained in the first weeks of experimentation with data obtained under the same experimental conditions six months later. The three *S*s were male undergraduates with normal vision who were paid for their services.

RESULTS

The data for individual *S*s are presented in minutes of arc per second in Table 1, and the logarithms of the means of these values are plotted in Fig. 2. The functions are all of the same general shape. Threshold velocity decreases with increase in

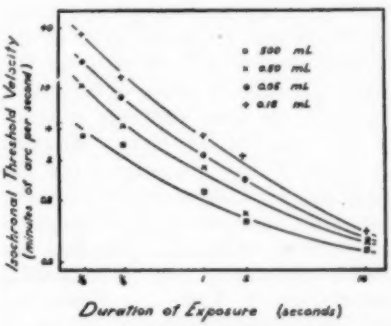

FIG. 3. Isochronal threshold velocity as a function of duration of exposure with luminance as parameter.

luminance, rapidly at first and then more slowly, before reaching a limiting value after which increase in luminance has little effect. The effect of exposure time is manifested primarily in the position of the curves on the threshold velocity axis. Increase in exposure time shifts the entire functions to progressively lower values. This effect can also be seen in Fig. 3 in which threshold velocity has been plotted as a function of exposure time with luminance values of 500, .5, .05, and .008 mL as parameters. These curves show more clearly the decrease in log threshold velocity with log exposure time and point to the more rapid decrease in this relationship at the lower luminance levels.

In addition to the effect of positioning the functions on the log threshold velocity axis, exposure time influences the shape of these functions. There is a tendency, with increase in exposure time, for the functions to cover a shorter range of velocities, i.e., the difference between thresholds obtained at the lowest and highest luminance decreases. This effect is most marked in going from the 2-sec. to the 16-sec. exposure time. There is also a tendency for the luminance value at which the curves level off to decrease with increased exposure time. The functions for the $\frac{1}{8}$- and 1-sec. exposures are decreasing even at the highest luminance level used, for the $\frac{1}{4}$-sec. duration it levels off at about 5 mL, for the 2-sec. exposure at about .5 mL, and for the 16-sec. duration at about .05 mL. In order to test the significance of these tendencies, the data were analyzed by the orthogonal polynomial technique described by Fisher and Yates (**4**). The results summarized in Table 2 indicate, for the orthogonal luminance values of .05, .5, 5, 50, and 500 mL, that the $\frac{1}{8}$- and $\frac{1}{4}$-sec. durations are fitted by both linear and quadratic components, the 2-sec. duration by a linear component, while no components are significant for the 1- and 16-sec. data.

TABLE 2

ANALYSIS BY ORTHOGONAL POLYNOMIALS OF ISOCHRONAL THRESHOLD VELOCITIES FOR VARIOUS LUMINANCE VALUES AND DURATIONS OF EXPOSURE

Exposure (Seconds)	Source	*df*	Mean Square	*F*
$\frac{1}{8}$	Linear	1	420.23	39.09***
	Quadratic	1	76.00	7.07*
	Cubic	1	1.08	
	Quartic	1	0.53	
	Within means	10	10.75	
	Total	14		
$\frac{1}{4}$	Linear	1	49.50	29.64***
	Quadratic	1	17.40	10.41**
	Cubic	1	3.47	2.08
	Quartic	1	0.04	
	Within means	10	1.67	
	Total	14		
1	Linear	1	2.902	4.05
	Quadratic	1	0.195	
	Cubic	1	0.006	
	Quartic	1	0.022	
	Within means	10	0.716	
	Total	14		
2	Linear	1	0.837	10.46**
	Quadratic	1	0.345	4.31
	Cubic	1	0.088	1.10
	Quartic	1	0.033	
	Within means	10	0.080	
	Total	14		
16	Linear	1	0.00108	
	Quadratic	1	0.00420	
	Cubic	1	0.00972	1.384
	Quartic	1	0.00000	
	Within means	10	0.00702	
	Total	14		

* .05, ** .01, *** .001 level of confidence.

DISCUSSION

Comparison of the present data with the results of previous investigators is difficult because of variations in the procedure for obtaining thresholds and the absence of specification of parameters such as velocity, displacement, exposure time, and luminance. However, to the extent that it is possible to make comparisons, most of the values found by

other investigators fall within the range of data in the present study. For example, values such as velocity thresholds of 1 min. of arc per sec. and displacement threshold of 20 sec. of arc (**5**) are of the same order of magnitude, after appropriate transformations, as some of the data reported here.

The most directly comparable studies in the literature are those of Rock (**14**) and of Brown and Conklin (**3**). Rock presented a moving grating test object to *S*s and determined the rate threshold with unlimited time of exposure. He found that as luminance was varied from .005 to 10 ft.-L. the velocity threshold decreased from .40 to .17 min. of arc per sec. This compares well with the range of the present data for the 16-sec. exposure, obtained for a slightly higher range of luminance values, of from .66 to .25 min. of arc per sec. Brown and Conklin determined threshold velocities for a dim (.026 mL) circular spot of light subtending 38 min. of visual angle and presented for exposure times from ½ to 16 sec. Their highest value of about 18 min. of arc per sec. for their ½-sec. exposure agrees with the results of the present experiment. However, their lowest value of 9 min. of arc per sec. is approximately thirty times higher than the data for comparable experimental conditions reported by Rock and in the present study. The main difference between the Brown and Conklin experiment as compared with the studies by Rock and the present investigator is with respect to the type of test objects employed. Brown and Conklin used a single circle of light, while Rock used a grating test object and the present apparatus employed a series of rectangles. It would be of considerable interest if the repeating pattern type of test object as opposed to the circle could account for the large differences in the thresholds obtained. A functional analysis of the relevant variables influencing motion perception, which would include variation in the configurations of the test objects, could be expected to provide information concerning this problem.

The general shape of the functions obtained in the present study, i.e., the increase in discriminatory performance with increase in luminance, rapidly at first and then more slowly, and finally approaching a limiting value, is familiar in the psychophysics of vision. Intensity discrimination, flicker, grating acuity (**10**), and vernier acuity (**1**) all exhibit similar relationships. To the extent that motion perception thresholds behave as these other measures of visual response, we might infer that similar mechanisms, probably photochemical, are critical to the perception of movement.[3] However, with longer durations of exposure, this function tends to become flat (see Fig. 2). A possible explanation is suggested by the distinction made by Exner (**15**) between the "direct" perception of motion and the "inference" of movement based on observation of change in position. The accuracy of localization of a target, provided it is seen, has been shown to be independent of luminance (**13**). It is to be expected, therefore, that motion perception, to the extent that it is based on "inference" of movement based on observation of change in location, would also become independent of the adapting luminance. This hypothesis is consistent with the present data which demonstrate a decreasing sensitivity of the isochronal threshold velocity to luminance changes as the exposure duration is increased.

This analysis suggests that photochemical mechanisms, while critical to motion perception for short durations of exposure, become less important as the duration is increased. Another indication of the diminished importance of photochemical events is seen in the systematic shift of the curves in Fig. 2 downward on the isochronal threshold velocity axis as the duration is increased. For other measures of visual discrimination such as the absolute threshold (**8, 12**), differential threshold (**7**), grating acuity

[3] Dr. Robert H. Brown has reported that the Bunsen-Roscoe law of photochemistry is applicable to the perception of motion for durations of exposure less than .1 sec. See reference **2**.

(6), and "span" of attention (11), increase in duration of exposure beyond a critical value of about .1 sec. has no effect upon the threshold. These results are interpreted as resulting from the critical importance of photochemical events with respect to these visual measures, the effective action of which is complete within the critical duration (9). In the present experiment, all of the durations employed, with the possible exception of the shortest value of .125 sec., are longer than this critical duration. The systematic decrease in the isochronal threshold velocity under these conditions is consistent with the hypothesis that movement perception for these durations becomes less dependent upon critical photochemical events and more dependent upon the "inference" of movement.

Summary

The threshold velocity for the perception of motion at a constant duration of exposure, designated the isochronal threshold velocity, is investigated for various luminance values and at five durations of exposure.

Threshold velocity decreases with increasing luminance, rapidly at first and then more slowly before reaching a limiting value. With increase in duration of exposure, the threshold velocity-luminance function becomes less sensitive to luminance changes, and the entire functions are shifted to lower velocity values.

The results are interpreted as reflecting the increased dependence of motion perception on the "inference" of movement at the longer durations of exposure and, under the same conditions, the decreased importance of "critical" photochemical events to the perception of movement.

References

1. Baker, K. Some variables influencing vernier acuity. *J. opt. Soc. Amer.*, 1949, **39**, 567–576.
2. Brown, R. H. The visual discrimination of velocity as a function of stimulus duration and luminance. *U. S. Naval Res. Lab. Rep.*, 1954, No. 4372.
3. Brown, R. H., & Conklin, J. E. The lower threshold of visible movement as a function of exposure time. *Amer. J. Psychol.*, 1954, **67**, 104–110.
4. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver & Boyd, 1943.
5. Graham, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 868–920.
6. Graham, C. H., & Cook, C. Visual acuity as a function of intensity and exposure time. *Amer. J. Psychol.*, 1937, **49**, 654–661.
7. Graham, C. H., & Kemp, E. H. Brightness discrimination as a function of the duration of the increment in intensity. *J. gen. Physiol.*, 1938, **21**, 635–650.
8. Graham, C. H., & Margaria, R. Area and the intensity-time relation in the peripheral retina. *Amer. J. Physiol.*, 1935, **113**, 299–305.
9. Hartline, H. K. Intensity and duration in the excitation of single photo-receptor units. *J. cell. comp. Physiol.*, 1934, **5**, 229–247.
10. Hecht, S. Rods, cones, and the chemical basis of vision. *Physiol. Rev.*, 1937, **17**, 239–283.
11. Hunter, W. S., & Sigler, M. The span of visual discrimination as a function of time and intensity of stimulation. *J. exp. Psychol.*, 1940, **26**, 160–179.
12. Karn, H. W. Area and the intensity-time relationship in the fovea. *J. gen. Psychol.*, 1936, **14**, 360–369.
13. Leibowitz, H., Angrist, N., & Grant, D. A. Radical localization of a single stimulus as a function of luminance and duration of exposure. *J. opt. Soc. Amer.*, in press.
14. Rock, M. Visual performance as a function of low photopic brightness levels. Dayton, O.: U. S. Air Force, Air Materiel Command, Wright-Patterson Air Force Base, 1950. (*AF Tech Rep.* No. 6013.)
15. Von Kries, J. In von Helmholtz, H., *Handbuch der physiologischen Optik.* Vol. III. (Trans. by J. P. C. Southall.) Opt. Soc. Amer., 1925. Pp. 270–273.

(Received May 24, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

DARK ADAPTATION AFTER INTERMITTENT AND CONTINUOUS PRE-EXPOSURES OF EQUAL DURATION [1]

F. A. MOTE AND ELEANOR R. ADAIR

University of Wisconsin

In a recent experiment (**3**) we studied the effect upon human dark adaptation of pre-exposure to continuous and two conditions of intermittent light. The two intermittent conditions were related to the continuous as follows: (*a*) For one of the intermittent conditions the durations of pre-exposure were the same as for the continuous, but the *intensities* were increased by the amounts necessary to equate the total quantity of light to that of the continuous condition. Comparisons could thus be made between a continuous pre-exposure of a given duration and intensity and intermittent pre-exposures of the same duration but higher intensities. No differences were found between the two modes of pre-exposure for durations of 1 and 3 min. and intensities varying over a range of 16 to 1. (*b*) For the other intermittent condition the intensities were the same as for the continuous, but the *durations* were increased by the amounts necessary to equate the total quantity of light. In this instance, comparisons could be made between a continuous pre-exposure of a given duration and intensity and intermittent pre-exposures of the same intensity but longer duration. In many cases when the duration was lengthened in order to compensate for the different amount of light time in the cycle of intermittence, a marked effect was obtained: for long durations the recovery of dark adaptation was retarded. It is well known that one of the effects of increasing pre-exposure duration, at a given intensity, is to slow down the rate of recovery of dark adaptation as compared with the recovery after the same intensity for a shorter duration. In the present experiment we gathered data for continuous pre-exposures equal in duration and intensity to the intermittent in order to compare the results under the two conditions.

[1] This research was supported by the Research Committee of the Graduate School of the University of Wisconsin from funds supplied by the Wisconsin Alumni Research Foundation.

Apparatus and Procedure

The apparatus was the same as that used in our previous experiment and as described by Mote and Riopelle (**4**). The procedure for measuring the threshold was as described by Mote, Riopelle, and Meyer (**5**). In the previous experiment (**3**) the intensities were 140, 281, and 562 mL and their light-dark ratios and durations were: $\frac{1}{3}$–$\frac{1}{3}$, 2 and 6 min.; $\frac{1}{4}$–$\frac{3}{4}$, 4 and 12 min.; $\frac{1}{8}$–$\frac{7}{8}$, 8 and 24 min.; and $\frac{1}{16}$–$\frac{15}{16}$, 16 and 48 min. In the present experiment we gave continuous pre-exposures 2, 4, 6, 8, 12, 16, 24, and 48 min. in duration to intensities of 140, 281, and 562 mL.

In all other respects the experiment was conducted as described previously (**3**). The intensities were controlled by inserting Wratten neutral tint fixed filters in the exposure beam. The retinal area exposed was 35° visual angle and the threshold testing area 5°. The pre-exposure stimulus was white, the testing stimulus violet. Both the pre-exposure and testing areas were imaged 7° from the fovea on the nasal portion of the right retina and all were made through a 2-mm. artificial pupil. The sessions were tabulated and their order of succession in the experiment randomly assigned. The authors served as *S*s and *E*s.

Results

A total of 48 dark-adaptation curves was obtained, one for each of

8 durations, for 3 intensities of pre-exposure, and for 2 *Ss*. The features of these data are similar to those found in other experiments carried out under similar conditions. The initial threshold values are all about the same and, since the intensities were all relatively low, there is little evidence of the discontinuity representing the transition from cone to rod adaptation. An inspection of all the data reveals that it is at the 562-mL intensity, and then only for durations of 8 min. and longer, that a rod-cone "break" is apparent in the curves. At all intensities the curves for the shortest duration have the steepest slope, the others all have about the same slope, and for these the slope is flatter than for the shortest duration. Data with features similar to these were obtained by Müller (**6**) and Wald and Clark (**7**) for durations of the order of 10–40 min.

In Fig. 1 and 2 are shown a number of pairs of curves for continuous and intermittent pre-exposures of the same duration. There are several points of difference between the curves of a pair. In every case the value of the initial threshold for the intermittent pre-exposure is lower than for the continuous; in most cases it is ten times lower. It is also seen that as dark adaptation proceeds most of the intermittent curves fall rapidly for 2 or 3 min. and then, rather suddenly, the rate of fall slows down. Whereas the threshold may have descended as much as two logarithmic units within the first 5 min., for some of the curves the descent during the following 10 min. may be no more than half a logarithmic unit. In most instances the thresholds keep dropping gradually, those for the longest duration reaching their final threshold level after the continuous curves for the same duration. Sometimes the retardation is so pronounced that the intermittent curve crosses over the

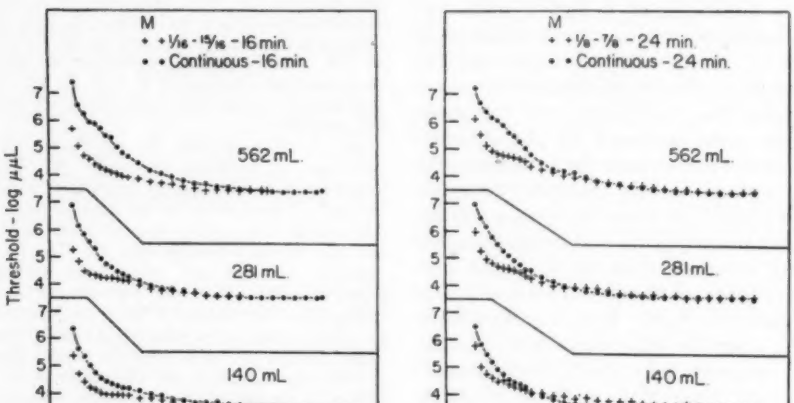

FIG. 1. Dark-adaptation curves obtained after intermittent and continuous pre-exposures of equal durations for three intensities for Subject M.

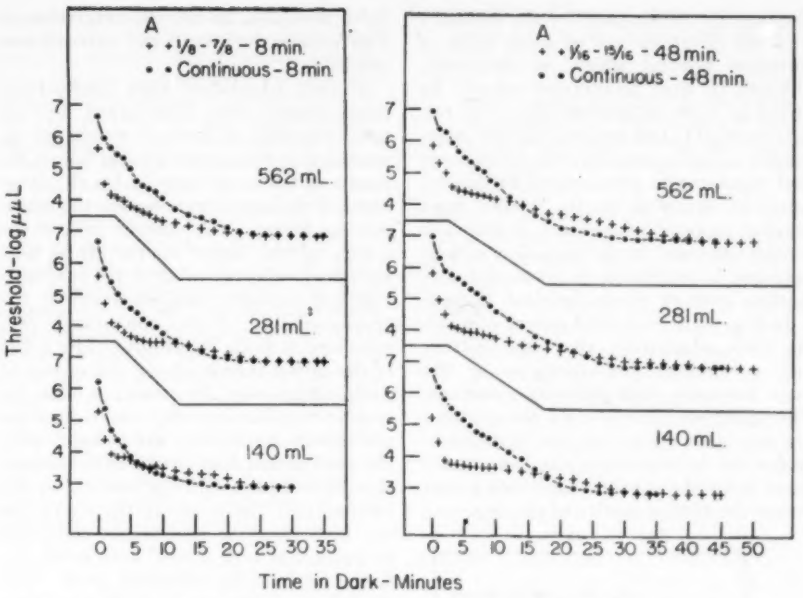

FIG. 2. Dark-adaptation curves obtained after intermittent and continuous pre-exposures of equal durations for three intensities for Subject A.

more rapidly falling continuous curve and stays above it until both reach the final level of dark adaptation.

The difference in the speed of recovery of dark adaptation for the two conditions of pre-exposure is all the more striking when we consider that the total quantity of pre-exposure light is always greater for the continuous condition. In order to keep the total quantity of pre-exposure the same for the intermittent conditions their durations were allotted according to the amount of light time in the light-dark ratios; consequently all those for the same intensity have the same quantity of pre-exposure. This was not the case for the continuous pre-exposures of the same intensity; for this condition the total quantity of light was determined by the duration. For example, the continuous light of 140-mL intensity whose duration was 16 min. provided 2240 mL-min. pre-exposure. The intermittent light of the same intensity was also presented for 16 min., but only for $\frac{1}{16}$ sec. in each second, since the pre-exposure was at the $\frac{1}{16}$–$\frac{15}{16}$ light-dark ratio; this gave a total of 140 mL-min. pre-exposure. How this difference in the quantity of pre-exposure light affects the rate of dark adaptation for the two conditions at the longest duration is shown in Fig. 2. Here it is seen that the intermittent curves require 45–50 min. to descend about three logarithmic units, while the continuous curves require 35–40 min. to descend about four logarithmic units. Yet the average intensity is only one-sixteenth as great for the intermittent pre-exposure as for the continuous.

DISCUSSION

The course of dark adaptation after pre-exposure to continuous light is pre-

dominantly determined by the intensity and the duration—up to some value of duration beyond which no detectable changes in dark adaptation occur. In terms of a photochemical theory of retinal events (1) this result is usually interpreted as indicating that the breakdown and regeneration processes of the photosensitive material in the retina have reached an equilibrium state, so that any further increase in the duration of pre-exposure is ineffective in promoting any further over-all photochemical change.

In Fig. 3 are presented curves comparing dark adaptation after intermittent and continuous pre-exposures of the same intensity, but different durations. The upper solid curve is for the continuous condition whose duration is the same as for the intermittent; the lower solid curve is for the continuous duration that makes the total quantity of pre-exposure light the same as for the intermittent. The crosses represent the intermittent condition.

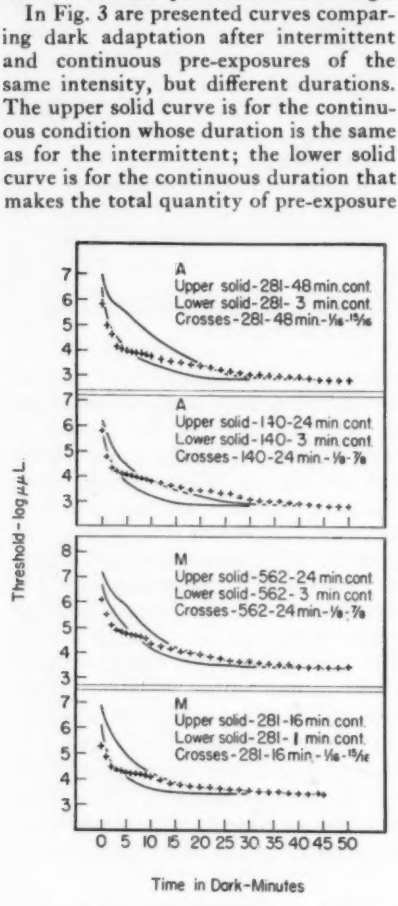
A
Upper solid - 281 - 48 min. cont.
Lower solid - 281 - 3 min. cont.
Crosses - 281 - 48 min. - 1/16 - 15/16
A
Upper solid - 140 - 24 min. cont.
Lower solid - 140 - 3 min. cont.
Crosses - 140 - 24 min. - 1/8 - 7/8
M
Upper solid - 562 - 24 min. cont.
Lower solid - 562 - 3 min. cont.
Crosses - 562 - 24 min. - 1/8 - 7/8
M
Upper solid - 281 - 16 min. cont.
Lower solid - 281 - 1 min. cont.
Crosses - 281 - 16 min. - 1/16 - 15/16
Threshold - log μμL.
0 5 10 15 20 25 30 35 40 45 50
Time in Dark-Minutes

FIG. 3. Curves comparing the course of dark adaptation after intermittent pre-exposures and two durations of continuous pre-exposure at the same intensity.

If dark adaptation after intermittent pre-exposure were determined by the total quantity of light, it would be expected that the crosses should follow the course of the lower curve. On the other hand, if the long duration were the determining factor, they should follow the course of the upper curve. It is seen that dark adaptation after the intermittent pre-exposure exhibits some of the characteristics of the continuous pre-exposure at both durations. The value of the initial threshold and the course of dark adaptation for about 5 min. is similar to that for the short-duration continuous condition. At about 5 min. the retardation begins and dark adaptation slows down, with the result that the intermittent curve meets the curve for the long-duration continuous condition and proceeds with it until both reach the final dark-adapted threshold level. For durations of the order of 16 min. and longer, dark adaptation after intermittent pre-exposure may be as retarded, or even more retarded, than after pre-exposure to continuous light of the same intensity and duration, despite the fact that the average intensity of pre-exposure for the intermittent condition is only a fraction of that for the continuous.

The photochemical hypothesis proposed by Wald (7), or some variant of it (2), has been successful in accounting for dark adaptation after different durations of continuous pre-exposure. However, until there is more information about the rates of breakdown and regeneration of rhodopsin for various conditions of pre-exposure, attempts to explain the phenomena of dark adaptation in photochemical terms will be incomplete. The data presented here show that the patterning of light and dark during pre-exposure can exert a marked influence upon the value of the initial threshold and the subsequent course of dark adaptation.

SUMMARY

Dark adaptation after continuous pre-exposures of 2, 4, 6, 8, 12, 16, 24, and 48 min.

was measured for two *Ss* and the results compared with dark-adaptation measurements after intermittent pre-exposures of the same duration. The rate of intermittence was 1 cycle/sec. and the ratios of light to dark in the cycle were $\frac{1}{2}$–$\frac{1}{2}$, $\frac{1}{4}$–$\frac{3}{4}$, $\frac{1}{8}$–$\frac{7}{8}$, and $\frac{1}{16}$–$\frac{15}{16}$. The intensities were 140, 281, and 562 mL. The total quantity of pre-exposure light for the continuous condition was always greater than for the intermittent condition of the same intensity and duration with which it was compared.

For the shorter durations, up to about 6 min., the curves of dark adaptation after both conditions of pre-exposure were similar. For longer durations it was found that after intermittent pre-exposure dark adaptation proceeded rapidly for about 5 min. and then slowed down. The retardation was so pronounced that the time required to reach the final dark-adapted threshold level was nearly always as long, and sometimes it was longer, than the time required after continuous pre-exposure, despite the fact that in some instances the quantity of pre-exposure light for the continuous condition was 16 times greater than for the intermittent.

References

1. Hecht, S. Rods, cones, and the chemical basis of vision. *Physiol. Rev.*, 1937, **17**, 239–290.
2. Jahn, T. L. Kinetics of visual dark adaptation. *J. opt. Soc. Amer.*, 1946, **36**, 659–665.
3. Mote, F. A., & Reed, E. C. The effect of high intensity and short duration *versus* low intensity and long duration of intermittent pre-exposure upon human dark adaptation. *J. opt. Soc. Amer.*, 1952, **42**, 521–525.
4. Mote, F. A , & Riopelle, A. J. The effect of varying the intensity and the duration of pre-exposure upon foveal dark adaptation. *J. gen. Physiol.*, 1951, **34**, 657–674.
5. Mote, F. A., Riopelle, A. J., & Meyer, D. R. The effect of intermittent preadapting light upon subsequent dark adaptation in the human eye. *J. opt. Soc. Amer.*, 1950, **40**, 584–588.
6. Müller, H. K. Über den Einfluss verschieden langer Vorbelichtung auf die Dunkeladaptation und auf die Fehlergrösse der Schwellenreizbestimmung wahrend der Dunkelanpassung. *Arch. Ophthal., Berlin*, 1931, **125**, 624–642.
7. Wald, G., & Clark, A. B. Visual adaptation and chemistry of the rods. *J. gen. Physiol.*, 1937, **21**, 93–105.

(Received May 26, 1954)

Journal of Experimental Psychology
Vol. 49, No. 3, 1955

A RETINAL EXCITATION GRADIENT IN A UNIFORM AREA OF STIMULATION[1]

LAWRENCE KRUGER AND JOHN R. BONAME

Yale University

Numerous investigators have shown that as the area of retinal illumination is increased less brightness is required to obtain a threshold effect in the human *S*. In an attempt to explain this relationship in terms of the neural properties of the retina, Adrian and Matthews (**2**) demonstrated that the proximity of four points of light used as stimuli determines the latency of the resultant optic nerve discharge in the eye of the conger eel. Using a similar stimulus arrangement and measuring the critical fusion frequency in human *S*s, Granit (**5**) found a facilitative interaction in the peripheral retina as the four points become closer, an effect which could not be demonstrated in the fovea. Further experiments (**6**) suggested that the interaction of these multiple stimuli in the peripheral retina was the result of synaptic reactions in the eye which lead to summation. Summation effects were more easily demonstrated in the peripheral retina, presumably because of the high density of rods and the "convergence" of numerous rods onto a single retinal ganglion cell in the region.

Extensive electrophysiological investigation has provided further evidence of neural interaction. Hartline (**7**) was able to demonstrate that a single element in the frog retina has a wide "receptive field" with a central maximum. The overlap of numerous receptive fields presumably accounts for areal summation. Using the invertebrate (Limulus) eye, which does not have a complex interconnected neural network, Graham (**3**) showed that the area-intensity relation is not demonstrable in the absence of a neural apparatus for summation. In a recent study of the mammalian retina, Kuffler (**8**) showed that photic excitation of adjacent retinal points can reproduce a facilitation of excitation in single retinal neural elements.

Graham, Brown, and Mote (**4**) proposed a theory that accounts for the inverse relationship between intensity and area of stimulation at threshold in terms of the interaction of different parts of a stimulated retinal area. Their formulation applies to areas larger than 20′, and stems from the idea that the excitation at any point within an area of stimulation is inversely related to its distance from the center of stimulation. This means, that "when an area of the retina is illuminated, the area does not exhibit a uniform degree of excitation in all the associated nerve fibers. Rather, there seems to be a gradient of effect from the margin inward, the smallest degree of stimulation occurring in the outermost zones, while the greatest effect is in the center" (**4**, p. 566). The present study is an attempt to determine experimentally whether the gradient of excitability proposed in this hypothesis is found when a small light stimulus is superimposed at different places on a larger dimly lighted circular area.

[1] The authors wish to express their thanks to Dr. Lloyd H. Beck for his help in designing the apparatus, and to Drs. Walter R. Miles and C. H. Graham for their advice and encouragement; also to Lester Reiss, who served as *S*.

METHOD

The "excitability" of a uniformly illuminated retinal area was determined by measuring the threshold for a "point source" (a small area of light) superimposed upon different regions of a large, circular, uniform stimulus area projected by fixation upon a region of the peripheral retina with uniform sensitivity.

All observations were performed with the dark-adapted eye (30 min. in total darkness before each experiment). The *S* was seated in a chair with a headrest. All determinations were made in the peripheral retina centered 18°45′ to the left of the fixation point for the right eye, a region chosen for obtaining the most uniform population of receptors (**10, 11**). Fixation was maintained by means of a small red foveal fixation cross held at a constant illumination just above threshold. All determinations were made with the right eye so that the stimulus always appeared on the temporal side of the retina.

Apparatus.—The arrangement of the apparatus is shown in Fig. 1. The large circular area for excitation was always viewed at 18°45′ in the temporal retina, and subtended a visual angle of 3.71° of arc. This area consisted of an opal glass diffusing surface with a cardboard mask placed in front of it, with an aperture of 4.5 in. The opal surface was illuminated from behind by a 150-w. projection lamp powered by a stabilized 110-v. DC line. Wratten neutral density filters were used to vary luminous intensity. The large circular area was reflected off the surface of a blue plate glass filter set at 45° to the opal glass. This arrangement provides a correction of color temperature (to 3120° K.) as well as a surface which can both transmit and reflect two independent photic stimuli. The point source consisted of a small square aperture (1.4′ visual angle) placed in front of a light-tight box which contained a 25-w. bulb in a diffusing surface, also operated on the 110-v. DC line. The large excitation area is nearly 20,000 times as large as the small "test" light for which thresholds were obtained.

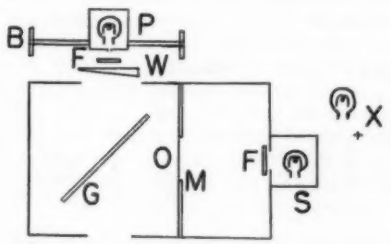

FIG. 1. The apparatus: *B*—optical bench; *F*—Wratten neutral density filters; *G*—blue plate glass for reflection of excitation source; *M*—cardboard circular mask aperture; *O*—opal glass diffusing surface; *P*—point source box containing lamp, diffusing surface, and aperture; *S*—excitation light source; *W*—logarithmic neutral density wedge; *X*—red foveal fixation cross.

The box containing the point source was mounted on an optical bench which was calibrated for the distance of the projected point source from the center of the large stimulus area. The point source was seen directly through the same blue filter which reflected the large area and could thereby be superimposed on the image of the latter. The intensity of the point source was controlled by a series of Wratten neutral density filters and a single logarithmic wedge. Balancing the wedge is not necessary for a small point. To avoid problems of lens accommodation the three lights used (fixation cross, large area, and point source) were kept at approximately the same distance from *S*.

Procedure.—The procedure maintained throughout the experiment was to seat *S* with the head in fixed position. The *S* was then asked to fixate the red fixation cross. After 30 min. of dark adaptation, the threshold for the large area was obtained for two ascending and two descending series. The average value was chosen for the intensity of the large excitation area and was maintained constant throughout each experimental session. The point source was set at one of the four points ranging from one periphery to the other along the horizontal plane within the large area, and a series of four ascending and four descending thresholds was obtained at each position of the point source. The duration of the flash was approximately 2 sec.; this was controlled by a hand-operated shutter. Adequate time was allowed between exposures for recovery of adaptation. The point source was then moved to another of the four positions and the same procedure was followed until 32 threshold determinations were obtained for each sitting. The positions were presented in different sequences on successive days. The observations were repeated on four days for each *S*. In order to determine the threshold for the point source alone, the same procedure was followed for four days, without using the large excitation light. Two trained *S*s were used in this experiment. In order to obviate guessing, "blanks" were presented at random intervals. Before a threshold determination was recorded, *S* was required to make two successive consistent responses.

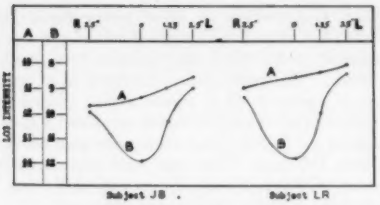

FIG. 2. Variation of the scotopic threshold for a point source presented in a homogeneous region of the peripheral retina when presented in the absence of other stimulation (A) and when superimposed on a large uniform circular area illuminating the same retinal location (B). The numbers at the top represent distance from the center of the large circular area toward the nasal (R) and temporal (L) borders. The total test area is approximately 3.7° of visual angle, which allowed 5 in. of movement on the optical bench scale. Each point represents the mean of 32 threshold determinations.

RESULTS

In order to determine whether the threshold for the point source varies at different positions within a large uniformly illuminated area, it was first necessary to determine the threshold at the same locations in the absence of the large excitation light as a control. Examination of Curve A in Fig. 2 shows that the region of the retina chosen is approximately uniform for both *Ss* and essentially confirms previous reports on the uniform sensitivity of the retina at approximately 20° (**10, 11**).

The failure to obtain a horizontal line (signifying uniform sensitivity) and the increased sensitivity at the nasal border on the retina appears to be statistically reliable (Table 1). The use of a smaller stimulus area in this study may explain the deviation from uniform sensitivity in this region of the retina demonstrated in Sloan's (**11**) study.

When the large excitation area was illuminated, the four points studied were superimposed on the horizontal meridian at the same retinal locations shown in Curve A, which serves as the control. The positions of the point source chosen superimposed at the center, at 2 mm. from each border, and at one-half the radius of the excitation area. Curve B (Fig. 2)

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE

Source	*df*	Subject JR		Subject LR	
		Mean Square	*F*	Mean Square	*F*
Control					
Days	3	5.25	63.25***	.233	4.46*
Positions	3	.94	11.33**	.593	11.36**
Residual (error)	9	.083		.052	
Total	15				
With Excitation Light					
Days	3	4.64	24.04***	2.34	6.41*
Positions	3	5.55	28.76***	7.95	21.78***
Residual (error)	9	.193		.365	
Total	15				

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.
*** Significant at .1% level of confidence.

shows the variation in threshold for the point source at different positions when the excitation area is simultaneously presented. Both *Ss* clearly show a statistically significant (Table 1) alteration in the threshold of the point source at each position. The threshold was found to be lowest in the center of the large area of illumination, and then rises as distance from the center increases, with highest thresholds occurring at the borders. The difference in sensitivity between the nasal and temporal borders is probably explicable in terms of the control data in Curve A. It is clear from a comparison of Curves A and B that a gradient of excitability can be detected in a homogeneously illuminated retinal area which is otherwise uniformly sensitive.

Discussion

The present experiment demonstrates the existence of a gradient of excitability in a homogeneous retinal region when uniformly illuminated. It has also been shown that the region of maximum facilitation is in the center of the area of illumination. The degree of facilitation, as measured by the threshold for a point source, is a function of the distance from the central maximum point, with minimal effect at the borders. Abney (1) arrived at a similar conclusion in determining the extinction thresholds for areas of different size. He noted that the center of an area, which is presumably the "recipient of the greatest amount of radiated stimuli," was the last to disappear when the luminous intensity was decreased.

These findings are in keeping with theoretical accounts which explain the decreased threshold with large stimulus areas in terms of neural interaction (4, 5) rather than a random distribution of elements with varying thresholds (12). Further evidence of neural interaction in the area-intensity relation can be found in Riopelle's experiment (9) which shows that increasing the dispersion of discontinuous stimuli of equal total area raises the threshold. The increased sensitivity of the elements in the center of the stimulated area is presumably attributable to the summation of neural events in the synaptic layers of the retina demonstrated electrophysiologically (2, 7, 8). Graham, Brown, and Mote's (4) prediction of a gradient of excitability with maximal facilitation of the centrally placed neural elements is supported by the present findings. The threshold sensitivity of a "point" in a uniformly illuminated retinal patch, is indeed a function of the distance from the central point of maximal neural convergence.

Summary

A method for testing excitation within a uniformly illuminated area is described, and intensity thresholds were obtained for a point source at four different horizontal positions within a large area in the dark-adapted human eye for the peripheral retina.

A consistent relationship is shown between the intensity threshold and the distance from the center of the large area, for the point source, the threshold being lowest in the center of excitation. It is believed that the present experiment provides direct confirmation of Graham, Brown, and Mote's (4) hypothesis explaining the area-intensity relation in terms of neural convergence.

References

1. Abney, W. *Researches in colour vision.* London: Longmans, Green, 1913.
2. Adrian, E. D., & Matthews, R. The action of light on the eye: III. Interaction of retinal neurones. *J. Physiol.*, 1928, **65**, 273–298.
3. Graham, C. H. The relation of nerve response and retinal potential to number of sense cells illuminated in an eye lacking lateral connections. *J. cell. comp. Physiol.*, 1932, **2**, 295–310.
4. Graham, C. H., Brown, R. H., & Mote, F. A. The relation of size of stimulus and intensity in the human eye: I. Intensity thresholds for white light. *J. exp. Psychol.*, 1939, **24**, 555–573.
5. Granit, R. Comparative studies on the peripheral and central retina: I. On interaction between distant areas in the

human eye. *Amer. J. Physiol.*, 1930, **94**, 41–50.

6. Granit, R., & Harper, P. Comparative studies on the peripheral and central retina: II. Synaptic reactions in the eye. *Amer. J. Physiol.*, 1930, **95**, 211–228.
7. Hartline, H. K. Nerve messages in the fibers of the visual pathway. *J. opt. Soc. Amer.*, 1940, **37**, 546–554.
8. Kuffler, S. W. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.*, 1953, **16**, 37–68.
9. Riopelle, A. J. Dispersion of stimulus-areas and the scotopic threshold. *Amer. J. Psychol.*, 1951, **64**, 380–390.
10. Riopelle, A. J., & Hake, H. W. Area-intensity relations in scotopic vision using annular stimuli. *J. exp. Psychol.*, 1951, **42**, 54–58.
11. Sloan, L. L. Rate of dark adaptation and regional threshold gradient of the dark adapted eye. Physiologic and clinical studies. *Amer. J. Ophthal.*, 1947, **30**, 705–720.
12. Wald, G. Area and visual threshold. *J. gen. Physiol.*, 1938, **21**, 269–287.

(Received May 28, 1954)